PIAAC Germany 2012: technical report

Zabal, Anouk; Martin, Silke; Massing, Natascha; Ackermann, Daniela; Helmschrott, Susanne; Barkow, Ingo; Rammstedt, Beatrice

Veröffentlichungsversion / Published Version
Monographie / monograph

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Anouk Zabal, Silke Martin,
Natascha Massing, Daniela Ackermann,
Susanne Helmschrott, Ingo Barkow,
Beatrice Rammstedt

PIAAC Germany 2012: Technical Report

COMMISSIONED BY

gesis
Leibniz Institute
for the Social Sciences

The Federal Ministry
of Education
and Research

WAXMANN

Anouk Zabal, Silke Martin, Natascha Massing,
Daniela Ackermann, Susanne Helmschrott,
Ingo Barkow, Beatrice Rammstedt

PIAAC Germany 2012: Technical Report

Waxmann 2014
Münster • New York

Commissioned by:

The Federal Ministry
of Education
and Research

With support from:

Federal Ministry
of Labour and Social Affairs

Acknowledgement: Our sincere thanks to Jana Kaiser for proofreading our manuscript.

Bibliographic information published by die Deutsche Nationalbibliothek
Die Deutsche Nationalbibliothek lists this publication in the
Deutsche Nationalbibliographie; detailed bibliographic
data is available in the Internet at http://dnb.d-nb.de

ISBN 978-3-8309-3113-3

www.waxmann.com
info@waxmann.com

Cover: Anne Breitenbach, Münster
Coverphoto: © stuart – fotolia.com
Typesetting: Stoddart Satz- und Layoutservice, Münster
Print: Mediaprint, Paderborn

Printed on age-resistant paper,
acid-free according to ISO 9706

Printed in Germany

Contents

1. Introduction

The *Programme for the International Assessment of Adult Competencies* (PIAAC) is a large-scale initiative of the *Organization for Economic Cooperation and Development* (OECD) that aims at assessing key adult competencies considered important for individual and societal success. PIAAC is a cooperative undertaking of the participating OECD countries, steered by a Board of Participating Countries and supported by the OECD Secretariat. The OECD contracted an international Consortium for the overall international project management of the first PIAAC survey, sometimes also referred to as the OECD Survey of Adult Skills (OECD, 2013a). This international Consortium was led by *Educational Testing Service* (ETS; USA), in cooperation with *cApStAn* (Belgium), *Centre de Recherche Public Henri Tudor* (CRP; Luxembourg), the *German Institute for International Educational Research* (DIPF; Germany), *GESIS – Leibniz Institute for the Social Sciences* (GESIS; Germany), the *International Association for the Evaluation of Educational Achievement – Data Processing and Research Center* (IEA DPC; Germany), the *Research Centre for Education and the Labour Market* (ROA; the Netherlands), and *Westat* (USA). The international Consortium was responsible for the overall design, coordination, and international implementation of PIAAC, and invested exceptional efforts into creating sound scientific and methodological groundwork for all the phases of the PIAAC implementation. A Technical Advisory Board advised the international Consortium in its work. Furthermore, international expert groups were involved in the development of the survey instruments.

In Germany, PIAAC was initiated and funded by the Federal Ministry of Education and Research with the participation of the Federal Ministry of Labor and Social Affairs. GESIS was contracted by the Federal Ministry of Education and Research to carry out the first round of PIAAC in Germany. A team of up to eight researchers was responsible for all aspects of the implementation of the PIAAC survey in Germany. This group formed the German National Center for PIAAC, which is a part of the Department of Survey Design and Methodology at GESIS. Thus, the German National Center was well placed to receive expert methodological advice from other GESIS professionals. Further support was provided by national content experts. A national Scientific Advisory Board, which included renowned scientists and government representatives, offered strategic advice and assistance to the National Center during the various project phases. Three national organizations were sub-contracted for specific work packages: (a) *TNS Infratest Sozialforschung,* Munich (TNS Infratest), the national survey organization, carried out the data collection; (b) DIPF was responsible for the national IT coordination;[1] and (c) IEA DPC conducted a number of coding and scoring activities. Figure 1.1 illustrates the structure of this national organization.

A list of persons who contributed to making the German PIAAC survey a success can be found at the end of this report. We would like to acknowledge and thank all the organi-

1 They also led the international technical software development.

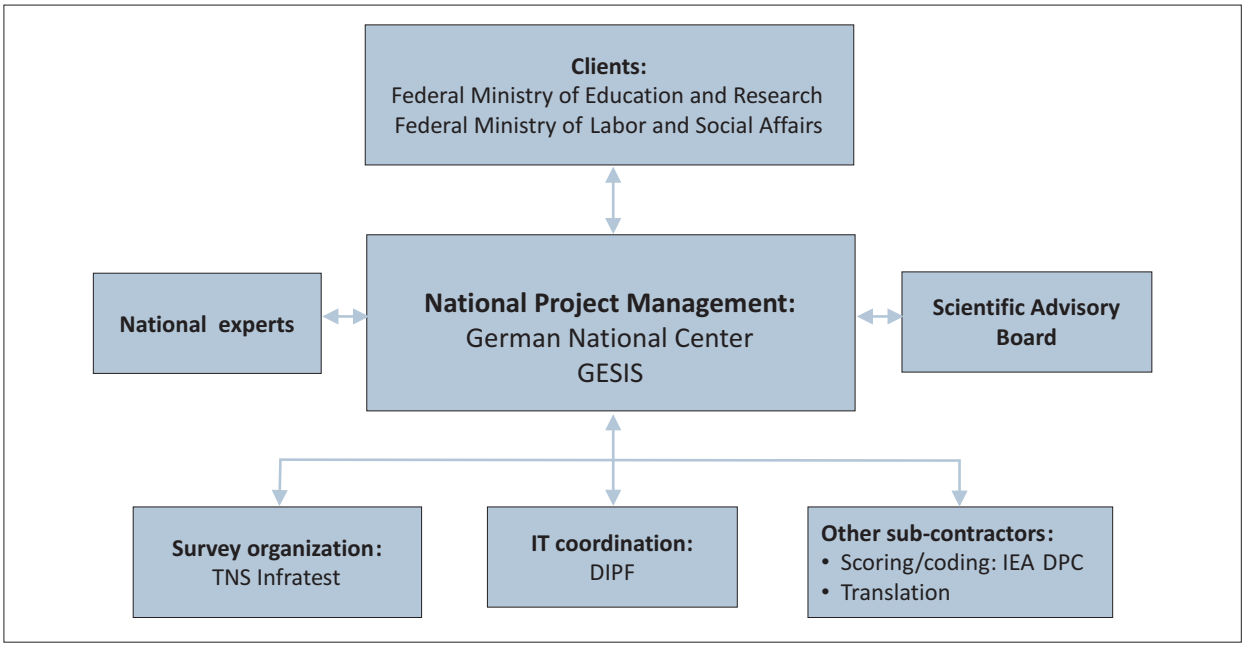


Figure 1.1. Organization structure of PIAAC in Germany

zations and individuals (including those who remained unnamed) for their valuable work, advice, and support. Furthermore, we would like to express our sincere gratitude to the German respondents for giving us their time and participating in our survey.

About PIAAC

PIAAC is intended to be a multi-cycle program for the assessment of adult competencies. This technical report refers to the implementation of the first round of the first cycle of PIAAC in Germany. This first round took place between 2008 and 2013. The second round of PIAAC, with nine additional countries, is currently taking place. A third round may come up in the near future. The second cycle of PIAAC is expected to start in 2018.

The data collection for the first round of PIAAC was carried out in 24 countries, including Germany, in 2011/2012. Key competencies of the adult population (ages 16 to 65) in the domains *literacy*, *numeracy*, and *problem solving in technology-rich environments* were assessed. This cognitive assessment was supplemented with a questionnaire that collected a wide variety of background information, including those related to demographic, social, educational, and economic variables. A probability-based sample representative of the target adult population was drawn in all countries. The one and a half to two hour interview generally took place at the respondents' homes. Respondents first answered questions from the background questionnaire, and then worked on the cognitive assessment. The assessment was per default on the computer; however, persons unfamiliar with computers were administered the assessment on paper.

International results from PIAAC 2012 (referred to as PIAAC in this report) were published by the OECD (2013a) in October 2013. At the same time, and as in many other countries, Germany released a national report focusing on German results (Rammstedt, 2013). The first PIAAC results have been received with interest by governments, the media, scientific communities, and the general public.

Key facts about the international PIAAC implementation, as well as the key specifics of the German implementation, are given in Box 1.1 (for international information see OECD, 2013a, 2013b, 2013c). In addition, some key facts regarding the field test that was carried out in all participating countries in preparation of the main survey are provided in Box 1.2.

Box 1.1: Key Facts About PIAAC

International Implementation

- Countries participating in PIAAC Round 1:
  - OECD countries: Australia, Austria, Canada, Czech Republic, Denmark, England/Northern Ireland (UK), Estonia, Finland, Flanders (Belgium), France, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, Slovak Republic, Spain, Sweden, United States
  - Non-OECD countries: Cyprus and Russian Federation
- Assessment domains: literacy, reading components (international option; implemented in all countries with the exception of Finland, France, Japan, and Russian Federation), numeracy, problem solving in technology-rich environments (international option; implemented in all countries with the exception of Cyprus, France, Italy, and Spain)
- Background questionnaire: demographic and background information, education and training, questions related to work
- Target population: non-institutionalized adults between 16 and 65 years of age who resided in the country during data collection, regardless of nationality, residential status, or language skills
- Sample: probability-based sample, representative of target population in country
- Target response rate: goal of 70%, at least 50%; for response rates below 70%, evidence was to be provided that there is either no or only limited nonresponse and undercoverage bias
- Minimum sample size: 5 000 completed cases (including problem solving in technology-rich environments) or 4 500 cases (excluding problem solving in technology-rich environments) per country or assessment language
- Data collection period: August 2011 to March 2012
- Interview administration: usually in respondent's home; in official national language(s); background questionnaire administered as a computer-assisted personal interview by trained interviewers; assessment self-administered by respondent (per default computer-based with an optional paper-based version), monitored by interviewer
- Interview duration: no timing restrictions; on average 80 to 95 minutes (background questionnaire between 30 and 45 minutes, assessment approximately 50 minutes)
- Total realized sample size: more than 160 000 respondents across all countries

Implementation in Germany

- Assessment domains: literacy, reading components, numeracy, problem solving in technology-rich environments (all optional domains were included); administered in German
- Sample: registry-based, two-stage stratified and clustered sampling design with selection of municipalities as primary sampling units or clusters at the first stage, and selection of individuals at the second stage
  - 320 sample points in 277 selected municipalities
  - 10 240 target persons (gross sample size)
- Survey organization: TNS Infratest
- Data collection period: 1 August 2011 to 31 March 2012
- Number of interviewers: 129
- Interview duration: on average 1 hour and 40 minutes (background questionnaire approximately 40 minutes, assessment approximately 60 minutes)
- Realized sample size: 5 465 respondents
- Achieved response rate (according to PIAAC definition; design weighted): 55%

Box 1.2: Key Facts About the PIAAC Field Test

International Implementation

- Key objectives of the field test:
  - Dress rehearsal for the main survey
  - Test accuracy, comparability, and timing of survey instruments
  - Test functioning of new computer delivery platform developed for PIAAC
  - Examine scaling procedures and evaluate psychometric properties of items and scales; for assessment, evaluate equivalence of computer and paper modes as well as linking to previous surveys IALS and ALL
  - Select final instrumentation for main survey, based on analyses of field test data
  - Check sampling, training, and survey operation procedures and materials
- Number of participating countries: 26; all of these countries, except Chile and Portugal, also completed the main survey
- Instruments: background questionnaire administered as a computer-assisted personal interview; assessment domains: literacy, reading components (optional), numeracy, problem solving in technology-rich environments (optional); computer-based and paper-based assessment
- Target population: non-institutionalized adults between 16 and 65 years of age who resided in the country during data collection, regardless of nationality, residential status, or language skills
- Sampling specifications: implementation of a probability-based random sample or convenience sample
- Minimum sample size: 1 500 completed cases per country or assessment language
- Data collection period: April to June 2010

Implementation in Germany

- Assessment domains: literacy, reading components, numeracy, problem solving in technology-rich environments (all optional domains were included); administered in German
- Sample: registry-based sample (three-stage stratified and clustered sample design); supplemented by a quota sample; sampling design was restricted to five federal states
- Survey organization: TNS Infratest
- Data collection period: 6 April 2010 to 30 June 2010
- Achieved sample size: 1 585 cases (including 1 185 cases from the random sample)
- Fieldwork response rate (unweighted) for random sample: 34%

Overview of Objectives and Features of PIAAC

PIAAC aims at producing high-quality data to assess and analyze the distribution of key competencies in adult populations of different countries and, in this way, at providing policy-makers with an empirically based foundation that informs their considerations and strategies. By offering empirical evidence on the competency profiles of adults, PIAAC makes it possible to examine the impact of these skills[2] on social and economic outcomes, and to explore how education and training systems, as well as other factors, may contribute to building and maintaining these skills. Modern societies increasingly demand that adults constantly adjust to changing conditions and continue to expand and refine their skills throughout their lifetime. The cognitive competencies measured in PIAAC can be regarded as generic in that they are prerequisites for many lifelong learning activities, and they are

2 In the context of PIAAC, the terms "competency" and "skill" are used interchangeably.

also crucial in terms of rapidly transforming modern technologies. Thus, knowledge about the stock of key skills in the adult population assessed in PIAAC is not only important in identifying skill shortages, but also in determining training and re-training demands, as well as for developing, shaping, and evaluating possible interventions and policies that promote the optimal use of available skills and their further development. The comparative perspective allows the exploration of similarities and differences across countries and cultures and offers countries the opportunity to learn from each other.

PIAAC stands in the tradition of two prior international assessments of adults' competencies: The *International Adult Literacy Survey* (IALS; OECD & Statistics Canada, 2000) and the *Adult Literacy and Life Skills Survey* (ALL; Statistics Canada & OECD, 2005). These two surveys assessed foundation skills, including prose literacy, document literacy, numeracy, and problem solving. Building on these two surveys, PIAAC has extended the spectrum of domains and construct coverage, the instrument delivery, and refined the methodology.

The assessment of the key cognitive competencies—literacy, numeracy, and problem solving in technology-rich environments—is at the heart of the first cycle of PIAAC. As defined in PIAAC, literacy is more than the mere ability to read. It encompasses understanding and evaluating written texts, as well as engaging with written texts and using them to achieve personal goals, expand one's knowledge, and to participate in society (Jones et al., 2009). It is important to note that the PIAAC literacy construct does *not* include writing texts. The conceptual PIAAC framework for literacy was based on the literacy frameworks of the preceding surveys, IALS and ALL. However, in comparison to the previous surveys, PIAAC extended the construct in two important respects: It included digital reading (previously, only traditional print texts had been considered), and it extended the measurement of literacy at the lower levels with the assessment of the reading components (Sabatini & Bruce, 2009).[3]

Numeracy, as conceptualized in PIAAC—and which builds on the ALL numeracy framework—is also a broader construct than simply the use of basic arithmetical skills. It involves an adult's ability to manage and respond to mathematical demands in a variety of everyday situations and contexts by accessing, using, interpreting, and communicating mathematical information and ideas (Gal et al., 2009).

PIAAC is the first large-scale assessment to include the domain problem solving in technology-rich environments.[4] As defined by the conceptual framework, problem solving in technology-rich environments refers to the ability to perform practical tasks using digital technology, communication tools and networks to access, process, and evaluate information (Rouet et al., 2009). The definition specifically covers only problems with which people are confronted when using *information and communication technologies* (ICT) and which require the active use of ICT. Furthermore, the framework stresses the communicative aspect of problem solving in technology-rich environments. The construct underlying problem solving in technology-rich environments unites aspects of cognitive problem-solving skills with computer or ICT literacy skills. Although the general concept embraces the entire range of digital technologies, the first PIAAC cycle has restricted its focus to computers.

In addition to the cognitive assessment, PIAAC implemented a background questionnaire designed to obtain a broad range of contextual information, in order to explore the relationship between the assessed skills and various social, educational, and economic outcomes. Thus, the PIAAC questionnaire included a wide variety of questions on topics such as educational qualifications, continuing education and training, work experience, work

3 The reading components were an international assessment option that was implemented in Germany and in all but four of the participating countries.

4 The domain problem solving in technology-rich environments was an international assessment option that was implemented in Germany and in all but four of the participating countries.

status, income, and other personal background information. An innovative component of the background questionnaire aimed at complementing the direct assessment of competencies with a set of questions revolving around the use of cognitive and non-cognitive skills at work.[5] This new element in the background questionnaire enables the exploration of questions related to skill mismatch. As an additional novelty, PIAAC extended this self-report of skill use to everyday life.

Developing the measurement instruments for the three assessment domains and for the background questionnaire was the responsibility of the international Consortium. In this, it was guided by international expert groups for each domain, and received contributions and comments from the participating countries. For literacy and numeracy, the instruments overlapped with those of the previous surveys, IALS and ALL, and also extended them. All of the items for the domain problem solving in technology-rich environments were novel and specifically developed for PIAAC. During item development for the assessment, special care was taken to create items that reflected commonplace tasks embedded in everyday contexts, which would be appropriate for the different cultures and age ranges of the PIAAC survey, would provide a full coverage of the underlying construct, and which would yield items and scales with appropriate psychometric characteristics. Central policy objectives and analytical interests guided the development of the background questionnaire, and various approaches were undertaken to validate the PIAAC questionnaire items.

PIAAC is the first international large-scale assessment that conducted the entire interview, consisting of the background questionnaire and the cognitive assessment, on the computer. Whereas the administration of the background questionnaire as a computer-assisted personal interview is standard in many surveys, the computer-based delivery of the direct assessment is not. A novel technology platform and assorted tools, partly based on, but also extending previously available software, were implemented in PIAAC. Various challenges were involved in the endeavor of delivering the interview software completely on the computer, which were exacerbated by the very restrictive project timelines. For example, it was not only necessary to create a stable PIAAC platform that encompassed a variety of software components, but this software also had to be delivered to and equally functional for all countries in the corresponding national versions. Furthermore, for the direct assessment, and with a view to linking the domains literacy and numeracy with the prior surveys, IALS and ALL, which was an important objective in PIAAC, previously paper-based items had to be transformed for comparable delivery on the computer. In addition, the development of new items for PIAAC explicitly intended to benefit from new features made possible by the computer-based assessment. This also included implementing an adaptive algorithm in the computer-based test delivery.

In summary, PIAAC aimed for an enriched measurement through the inclusion of a new domain, the extension of constructs from IALS and ALL, in part rendered possible through the computer-based implementation, as well as a comprehensive and analytically rich background questionnaire.

International Quality Assurance and Control

PIAAC also strove for best practice in all phases of the survey life cycle. In order to obtain a high-quality database with valid, reliable, and comparable data and results for all participating countries, the international Consortium established an elaborate set of quality assurance and quality control measures. As part of these measures, a comprehensive set of international standards and guidelines was compiled (OECD, 2010b). These standards and guidelines represented generally acknowledged best practices for survey implementation and

5 These questions were based on the Job Requirement Approach (Felstead, Gallie, Green, & Zhou, 2007).

included: (a) ethics, (b) survey planning, (c) sample design and weighting, (d) survey instruments, (e) translation and adaptation, (f) information technology, (g) field management, (h) interviewer selection and training, (i) data collection, (j) data capture and file creation, and (k) confidentiality and data security. Countries were required to adhere to these international standards, although it was possible for countries to apply for justified deviations (in certain instances). Compliance with the international requirements was closely monitored by the international Consortium. Part of the quality control process required countries to fill out numerous forms that kept the international Consortium informed about all aspects of national implementation and progress. In addition, countries received training, extensive documentation, and participated in quality control phone calls with the Consortium. Furthermore, the PIAAC field test was an important measure to ensure the functioning and quality of the main survey. All participating countries had to carry out a field test in 2010; more details on this field test (both from the international as well as the German perspective) can be found in Annex A.1.

The Implementation of the PIAAC Survey in Germany

The German National Center started its work in 2009. Although the entire project work was carried out within a very tight timeline, the National Center made every possible effort not only to achieve the high international standards, but also to enhance these with appropriate national quality measures. The project had three milestones: The *field test* in 2010, the *main survey* in 2011/2012, and the *publication of results* in 2013. In the initial project phase, in preparation for the field test, activities focused on three work packages. The first consisted of the development of the German instrumentation. The international source versions were translated into German and adapted to fit the German context. Furthermore, the international PIAAC software containing the German PIAAC instruments (delivered by the international Consortium) was integrated with the German case management software and tested extensively. The second work package covered developing the national survey design for the field test as well as a preliminary version for the main survey. The third package specified national field test survey operations and prepared national data management activities. This included developing study materials and the interviewer manual, interviewer training, as well as building up a national database structure (based on and extending international master versions).

Following completion of the field test, the instrumentation, the survey design, procedures, as well as study and training materials were revised and modified for the main survey. The German fieldwork for the main survey was especially carefully planned and aimed at bringing together a comprehensive set of suitable measures to ensure an excellent quality of the data collection, to achieve as high response rates as possible, and to limit nonresponse bias. This was also important with respect to the very challenging PIAAC standards regarding response rates and data quality.

The fieldwork for the main survey was closely monitored by both the survey organization and the German National Center. Special care was taken in the validation of interviews during this phase. Following the main survey data collection, the German National Center carried out an elaborate set of data management activities. Furthermore, weighting variables were selected and preliminary nonresponse bias analyses were prepared. Weighting for the German data was subsequently carried out by the international Consortium, along with international data cleaning and scaling. The international Consortium released various versions of the German national database, which included weights and plausible values (derived through proficiency scaling), to the German National Center starting early 2013. Final nonresponse bias analyses were carried out with this database, together with auxiliary paradata and benchmark data.

In addition to the databases, the international Consortium provided preliminary data analysis tools, specifically tailored to the PIAAC data structure, for the production of national reports. These were used to prepare the national report of the German results. The above-mentioned tools, as well as the national data files, were released by the OECD in a final version for public use and simultaneously with their international report. The German National Center had previously defined a set of necessary confidentiality edits to protect the identity of respondents in the German PIAAC data set released by the OECD. To provide the scientific community with a richer data set, a first Scientific Use File with German PIAAC data was prepared and released in the spring of 2014.

This technical report describes how the PIAAC survey was conducted in Germany. It includes information on the international features of PIAAC, as appropriate, but focuses on the German specifics (see OECD, 2013c for a comprehensive overview of the PIAAC survey from an international perspective). The second chapter provides information on the instrumentation. The third describes sampling, fieldwork, weighting, and nonresponse bias analyses. The final chapter gives an overview of the data management, data processing, and data products. It ends with a brief evaluation of the overall data quality.

2. Instrumentation

Assessing the domains literacy, numeracy, and problem solving in technology-rich environments (PS-TRE) in the adult population across more than 20 countries was a major objective of PIAAC. In order to measure these key cognitive competencies, a *direct assessment* was developed. Furthermore, a background questionnaire was created to gain rich context information, in order to be able to analyze factors potentially related to competencies. The development of the instruments in PIAAC was the responsibility of the international Consortium, with guidance from international domain expert groups, and with contributions from the participating countries. The aim was to create comparable instruments through input harmonization. The participating countries were responsible for translating and adapting these instruments, to create equivalent and adequate instruments for their national context. The international PIAAC software allowed the administration of the background questionnaire as a *computer-assisted personal interview* (CAPI) as well as a computer-based delivery of the direct assessment on a new technology platform.

This chapter first describes the interview workflow and assessment design (Section 2.1). It then introduces the background questionnaire (Section 2.2) and the direct assessment (Section 2.3). After presenting an overview of the translation process (Section 2.4), some information on the technical aspects is provided (Section 2.5).

2.1 Interview Workflow and Assessment Design

This section gives a brief overview of the interview workflow and assessment design (a variant of matrix design) implemented for the PIAAC main survey, based on the descriptions given in Kirsch and Yamamoto (2013) and OECD (2011b, 2013b), and also provides national specifications where appropriate.

PIAAC was an interviewer-mediated survey consisting of a background questionnaire and a direct assessment. The administration of the entire interview was computer-based. The background questionnaire, containing a variety of questions about the respondent, was administrated by an interviewer as a CAPI. In Germany, the duration of the background questionnaire was approximately 40 minutes, which is comparable to the average international duration.[1] The direct assessment was administered immediately after the background questionnaire and was per default computer-based—this was one major innovation of PIAAC, compared to the previous large-scale assessments of adult skills, IALS and ALL. Respondents who were either not capable of doing the assessment on the computer (e.g., because they had never worked with a computer before), or not willing to do so, could work on the assessment on paper. During the assessment, respondents worked on their own and

1 According to the OECD (2013a), the average range of the duration of the background questionnaire across all countries was between 30 and 45 minutes.

without any time limit; interviewers monitored respondents' progress and provided them with work material, such as a calculator. For the computer branch, interviewers handed over their laptop to respondents, whereas, in the paper branch, interviewers administered paper booklets to the respondents. The average duration of the assessment in Germany was approximately 60 minutes.[2] Figure 2.1 shows a simplified scheme of the interview workflow.[3] Although the PIAAC interview contained many components and possible routings, the international software directed interviewers and respondents smoothly through their individual paths.

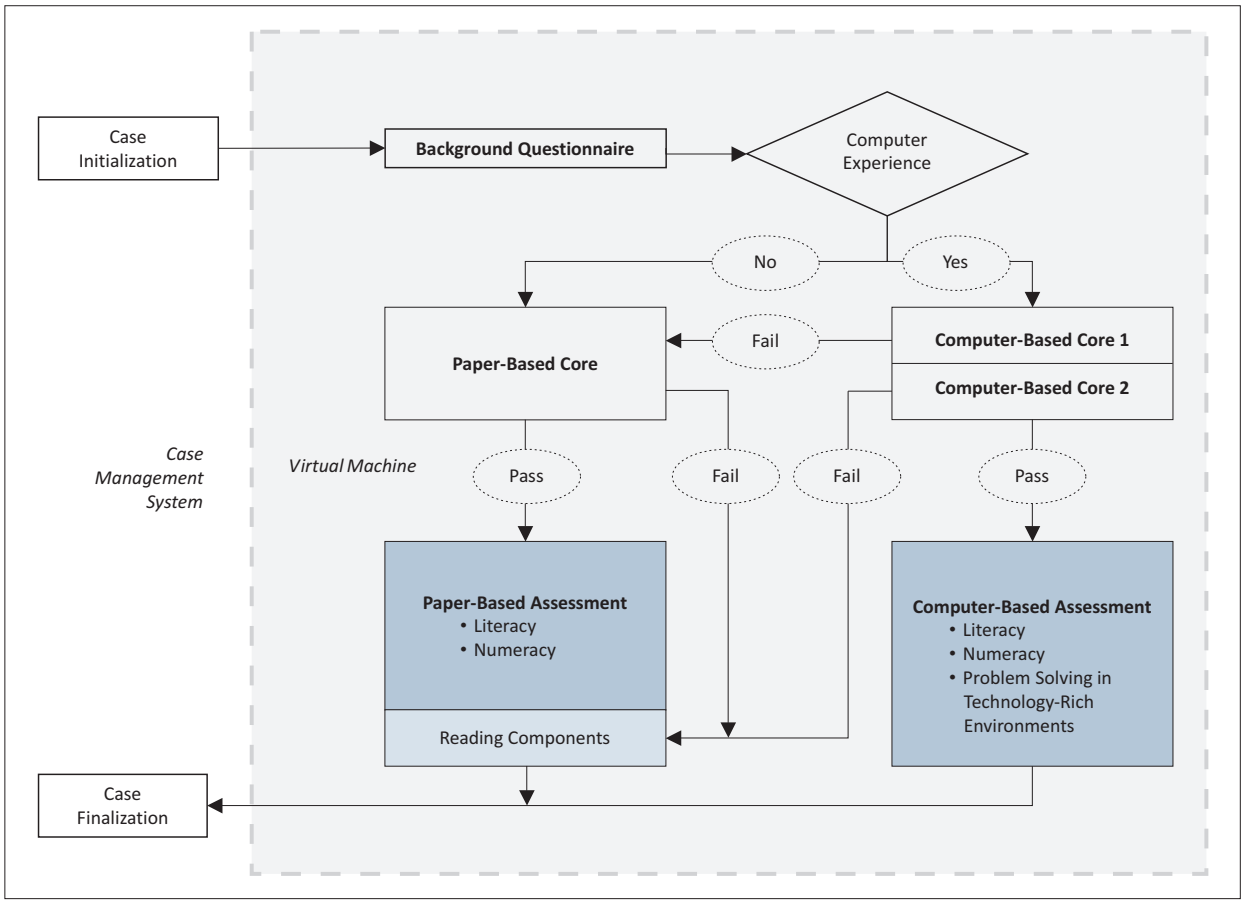


Figure 2.1. Scheme of PIAAC interview workflow

The PIAAC interview was implemented using special software and then assembled in a virtual machine (Jadoul, Plichart, Bogaerts, Henry, & Latour, 2013; Upsing et al., 2013a; also see Section 2.5). In Germany, the virtual machine was integrated into the case management system of the survey organization. Amongst other things, the case management system organized cases and transferred data from the interviewer laptop to the survey agency. After a case was initialized in the case management system, the interview began with the background questionnaire.

Some elements of the background questionnaire were important for the subsequent assessment. For example, respondents who reported having worked with a computer before (regardless of the amount of experience) were routed directly to the computer-based branch of the assessment, after completion of the background questionnaire. Respondents without any computer experience were routed directly to the paper-based assessment. In Germany,

2 This is somewhat longer than the international average length for the cognitive assessment of approximately 50 minutes (OECD, 2013a).
3 A more detailed workflow can be found in Annex A.2.1.

the interviewers were instructed not to offer the paper-based assessment as a choice. However, some respondents spontaneously refused to do the assessment on the computer and therefore received the paper-based assessment. In total, approximately 15%[4] of the respondents in Germany did the assessment on paper.

2.1.1 Computer Branch

The computer branch consisted of two main components: the *computer-based core* and the *computer-based assessment*. All components of the computer branch are shown in detail in Figure 2.2.

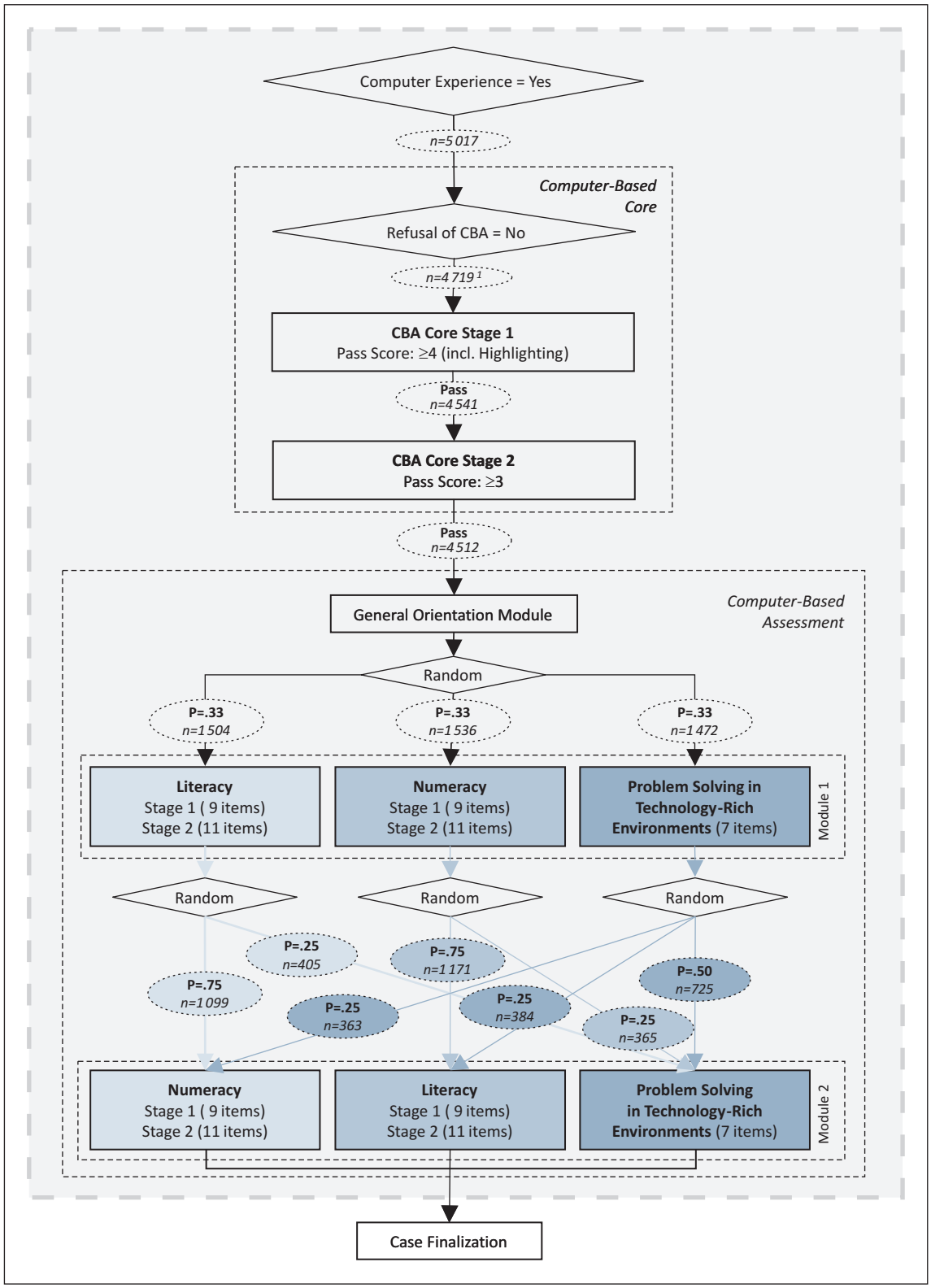
The computer branch started with the computer-based core, which included two stages: *CBA*[5] *Core Stage 1* and *CBA Core Stage 2*. Respondents first received CBA Core Stage 1, which consisted of six short tasks that tested the basic use of the computer mouse. The aim was to ensure that respondents were technically able to complete the computer-based assessment. Respondents needed to complete at least four tasks correctly to continue with the computer branch. One of these tasks required the respondent to *highlight* text, using the computer mouse. A correct response to this specific task was also a prerequisite to pass the CBA Core Stage 1, due to the fact that highlighting was essential for many of the computer-based assessment items. As indicated above, respondents who failed CBA Core Stage 1 were routed to the paper branch of the assessment. Respondents who passed CBA Core Stage 1 continued with CBA Core Stage 2. CBA Core Stage 2, consisting of basic literacy and numeracy items, determined whether respondents had sufficient basic literacy and numeracy skills to proceed with the computer-based assessment. To pass CBA Core Stage 2, at least three out of six items had to be answered correctly. Respondents who failed CBA Core Stage 2 were routed to the paper-based reading components.

As shown in Figure 2.2, the computer-based assessment started with a general orientation, which introduced respondents to the screen layout, response formats, and other general functionalities. The general orientation was followed by *Module 1*, which consisted either of literacy, numeracy, or PS-TRE items. Following the completion of Module 1, the respondent was routed to *Module 2*.

In Module 1, respondents were randomly allocated to a domain, with a probability of one third per domain. Module 1 started with a domain-specific orientation introducing the specific functionalities of the domain, such as spreadsheet functionalities for PS-TRE, highlighting functionalities for literacy, or entering numerical responses for numeracy. For literacy and numeracy, an adaptive algorithm was implemented to select item sets, i.e., task difficulty was adapted to a respondent's individual characteristics. Thus, the difficulty of the assessment items varied, depending on the information derived from the background questionnaire and the performance in previous parts of the computer branch. Adaptive testing enabled a deeper and more accurate assessment of respondents' ability level, while reducing respondents' burden.

Due to the adaptive assessment design implemented in PIAAC for literacy and numeracy, both Module 1 and Module 2 were organized in two stages. Each stage contained different sets of items with varying difficulty (testlets): three testlets in *Stage 1* (nine items) and four in *Stage 2* (eleven items). For both modules, only one testlet was administered per stage. Testlet allocation in Stage 1 depended on (a) respondents' education and mother tongue (both derived from the background questionnaire), (b) their performance in CBA Core Stage 2, and (c) a random element. The difficulty of the testlet selected in Stage 2 additionally depended on the respondents' performance in the testlet of Stage 1 of the same module.

4 Unweighted percentage (approximately 800 respondents)
5 CBA = computer-based assessment



Notes. Cp. Martin et al. (2013, p. 174) and OECD (2011b). There were twelve breakoffs in the computer-based assessment (five of them were technical breakoffs). CBA = computer-based assessment. P = allocation probability. *n* = number of cases.

[1] One technical breakoff is excluded.

Figure 2.2. Workflow of the computer branch

The assessment design for PS-TRE was not adaptive. Due to the length and complexity of the PS-TRE items, only two testlets were assembled. One testlet was administered in Module 1, and the other one in Module 2.

Analogous to Module 1, the domains in Module 2 were also randomly assigned. However, here the allocation probabilities depended on the domain that respondents had received previously in Module 1:

a) Respondents who had received literacy in the first module received numeracy, with a probability of 75%, and PS-TRE, with a probability of 25%, in the second module.
b) Respondents who had received numeracy in the first module received literacy, with a probability of 75%, and PS-TRE, with a probability of 25%, in the second module.
c) Respondents who had received PS-TRE in the first module received numeracy, with a probability of 25%, literacy with a probability of 25%, or PS-TRE, with a probability of 50%, in the second module.

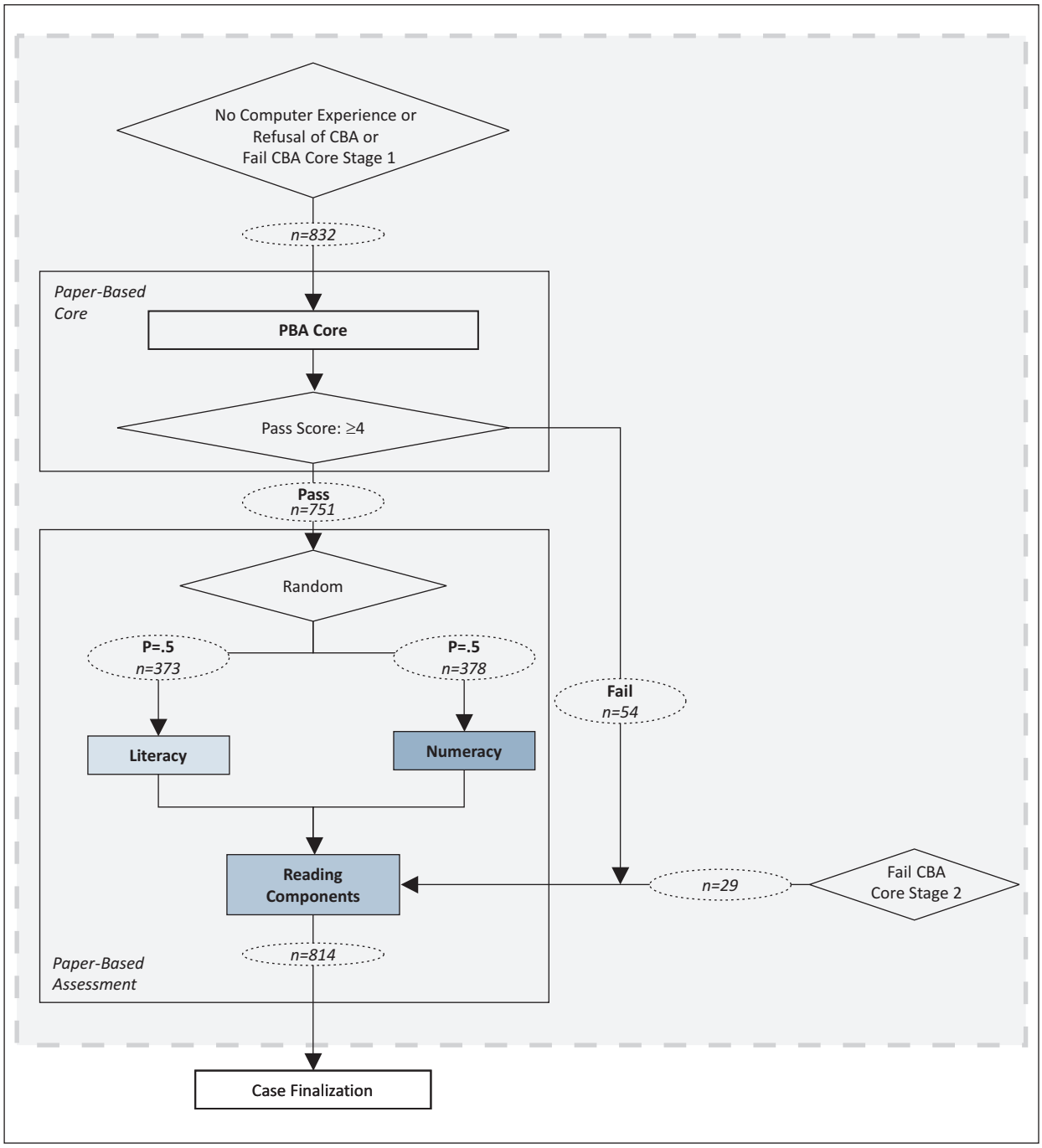
2.1.2 Paper Branch

The paper branch consisted of three main components: the *paper-based core*, the *paper-based assessment* for literacy or numeracy, and the *reading components*. Per definition, there was no paper-based assessment of PS-TRE. All components of the paper branch are shown in detail in Figure 2.3.

The paper branch started with the paper-based core (PBA[6] Core). This core consisted of the same items as in CBA Core Stage 2, plus two additional items. Since routing in the paper branch depended on respondents' performance in this core, these eight items were scored immediately by the interviewers. Respondents who answered a minimum of four items correctly subsequently received, at random, a booklet including either *literacy-only* or *numeracy-only* items (each booklet included 20 items). Respondents who failed the core skipped the main part of the paper assessment and immediately received the reading components booklet (which included 100 items). Respondents who passed the PBA Core first completed the literacy or numeracy booklet and then the reading components booklet. Thus, all respondents in the paper branch were administered the reading components.

After the PIAAC assessment was completed, interviewers answered questions about the interview setting and closed the case (*case finalization*).

6 PBA = paper-based assessment



Notes. Cp. Martin et al. (2013) and OECD (2011b). Breakoffs are excluded at the stage at which they occurred and the subsequent stages. Reported number of cases after the reading components module excludes respondents with missing reading components booklets. CBA = computer-based assessment. PBA = paper-based assessment. P = allocation probability. *n* = number of cases.

Figure 2.3. Workflow of the paper branch

2.2 Background Questionnaire

The background questionnaire for PIAAC was developed by the international Consortium (Allen et al., 2013), in cooperation with a Background Questionnaire Expert Group consisting of international experts, and with input from the countries. The two main goals during development were to create a background questionnaire that had the potential for analysis (especially in combination with the competency measures), and that created data that were internationally comparable. The international background questionnaire was developed in English and had to be adapted and translated in each country (see Section 2.4.3).

A framework for the background questionnaire that specified the analytical underpinnings was developed by the international Consortium in collaboration with the international expert group (OECD, 2011a). This framework provided a rationale for the constructs that were to be covered in the PIAAC background questionnaire. Based on this framework, the PIAAC background questionnaire was developed and covered topics such as sociodemographic information, education, and training, as well as questions related to work and the background of the respondent. As one of the innovations in PIAAC, some of the questions were taken from the so-called *Job Requirement Approach* (JRA; Felstead et al., 2007; OECD, 2011a). Its aim was to measure the relationship between competencies and the actual use of certain skills at work and, additionally, to offer indirect measures of skills that could not be included in the direct assessment. Concretely, the JRA asked respondents how often they carried out certain tasks in their job. Thus, it is possible, for example, to analyze how literacy practices at work relate to the literacy competencies measured in the assessment. The PIAAC background questionnaire extended this approach to activities in everyday life.

At the end of the development process, the international background questionnaire differentiated ten sections, as shown in Box 2.1.[7]

Box 2.1: Different Sections of the PIAAC Background Questionnaire

- *Section A – General information*: year of birth and gender
- *Section B – Education and training*: different educational qualifications (e.g., highest qualification or current education), detailed information on continuing education
- *Section C – Current status and work history*: respondent's occupational status (employed, unemployed, or out of the labor force), simplified employment history
- *Section D – Current work*: respondent's occupation, sector of industry, work satisfaction, working hours, salary/earnings
- *Section E – Recent work*: very similar to Section D but for last job; directed to respondents who were not working at the time of the interview: respondent's occupation, sector of industry, etc.
- *Section F – Skills used at work*: activities in the current or last job (e.g., cooperation with co-workers, planning activities, time management, physical work)
- *Section G – Skill use literacy, numeracy, and ICT at work*: reading, writing, numeracy and computer tasks carried out at work, computer experience
- *Section H – Skill use literacy, numeracy, and ICT in everyday life*: analogous to Section G, but related to everyday life
- *Section I – About yourself*: volunteering, social trust, political efficacy, and health
- *Section J – Background information*: children, country of birth, citizenship, parental education

7 The international background questionnaire can be accessed on the OECD PIAAC website (http://www.oecd.org/site/piaac/).

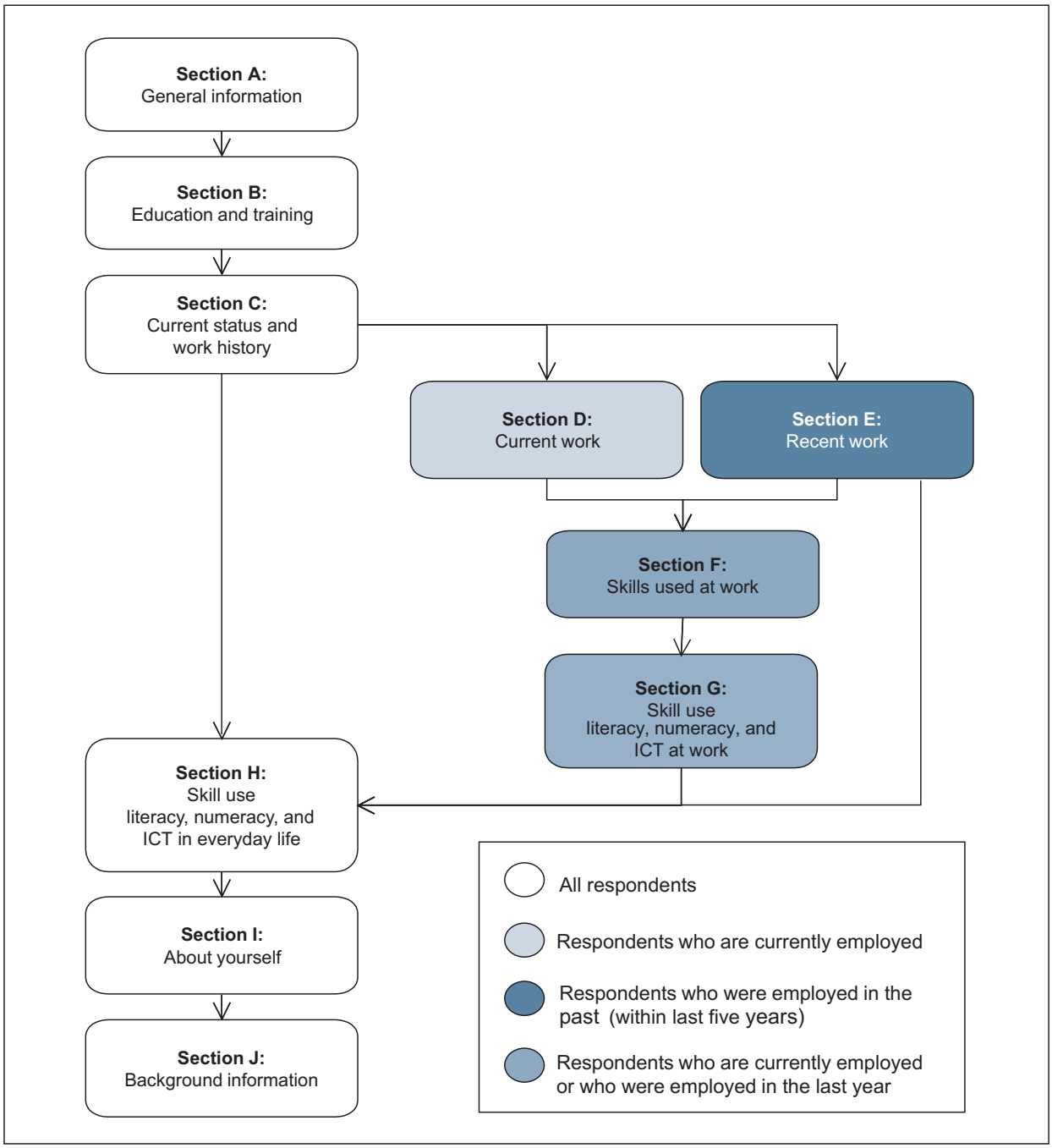


Figure 2.4. General routing of the background questionnaire

Because some of the topics covered in the background questionnaire were not relevant for all respondents, respondents received different sections of the background questionnaire, depending on their biography. The flow chart (Figure 2.4) shows which group of respondents was administered which sections of the background questionnaire. All respondents received Sections A to C as well as Sections H to J. However, only respondents currently working received Section D. Respondents with recent work experience (within the past five years) received Section E. Respondents currently working or with work experience during the last year also received Sections F and G.

The PIAAC background questionnaire aimed at being partly comparable to other surveys and also at implementing questions that had been previously tested. Therefore, many of the questions were taken from other surveys. Most of these questions originated from ALL. A number of questions were also taken from the Adult Education Survey (AES), the Labour Force Survey (LFS), and others. Moreover, there were some questions that were specifically designed for PIAAC. The exact sources for the background questionnaire questions can be found in the documented international background questionnaire, which can be accessed at the OECD PIAAC website referred to above.

The quality of the international background questionnaire was thoroughly validated. One important part of the validation strategy was the administration of an extended version of the final background questionnaire in the PIAAC field test. Question selection for the main survey questionnaire focused on (a) a content-based review of construct priorities at international and national levels and (b) item performance, as determined by the field test data analyses. Based on empirical analyses of these data, the most valid variables and indicators for proficiency, question sets that formed reliable scales, and questions with the lowest country-specific bias were selected.

2.2.1 German Adaptations and Extensions

The international background questionnaire had to be adapted by each country to accommodate for national systems and constraints and translated into its target language(s). The background questionnaire adaptation guidelines specified certain variables that had to be adapted to fit the national context. These adaptations referred to substantial changes, made necessary by differences in national systems (and not to strictly language-driven adaptations, which are a part of the translation process described in Section 2.4). Structural adaptations were implemented to reflect national differences, for example, in (a) educational and training institutions, (b) labor market institutions, (c) currencies and salary ranges for the measurement of income, and (d) origins of typical migrant populations. They involved changes to the question content, often with structural implications such as supplementary response categories, or even additional questions. In addition to such internationally prescribed adaptations, countries could apply for supplementary adaptations. However, these were to be restricted to an absolute minimum and national requests had to be well-justified and approved by the international Consortium. Furthermore, each country was allowed five extra minutes for national extension questions in the background questionnaire.

The international Consortium requested countries to document adaptations to the national background questionnaires in an Excel spreadsheet, which served as a communication and documentation tool. Box 2.2 gives an overview of the background questionnaire questions that were adapted for Germany.

Box 2.2: German Adaptations

a) Educational qualifications

In PIAAC, there were various questions relating to educational qualifications. These questions referred to a variety of contexts, such as the highest qualification, recent (last 12 months) and current education, and attempted, but not completed, educational qualification. Because of the nature of the educational system in Germany, two questions (in contrast to one question in the international background questionnaire) had to be asked to obtain the necessary information on educational background: the first related to general education (*allgemeinbildender Schulabschluss*) and the second to a professional/university qualification (*beruflicher Ausbildungsabschluss/Hochschulabschluss*). The response categories (i.e., different qualifications) for these questions were widely based on the Microcensus 2005 categories and the German demographic standards (Statistisches Bundesamt, 2004). National educational experts were also consulted. The fact that two questions and not one were needed to obtain the required information on educational qualifications also affected the routing of the German background questionnaire.

Furthermore, it was necessary to drop the international question on the number of qualifications obtained during the last year (variable name: B_Q04b).

b) Occupation

Countries were asked to implement occupation questions that reflected the best way, nationally, to obtain the level of detail required for coding into the *International Standard Classification of Occupations 2008* (*ISCO-08*; International Labour Office, 2012; see Section 4.2 for more details). The international background questionnaire used two questions here. In Germany, the wording was based on questions from various national surveys and the German demographic standards (Statistisches Bundesamt, 2004). The first question also included examples of potential responses, in order to show respondents the level of detail that was required.[1]

c) Industry

Similarly, the questions on industry had to be coded into the *International Standard Industrial Classification of All Economic Activities* (*ISIC*; United Nations Statistics Division, 2013a). The international background questionnaire specified two questions for industry that required an open response. Again, countries were asked to use the best available questions to obtain adequate information. The German questionnaire followed the international question structure. The wording for the first question was based on the question in the Socio-economic Panel (SOEP 2008), which also required an open response. The second question produced the same level of detail as the international background questionnaire.

d) Salary

The international background questionnaire included questions on the respondents' salaries (or wages/earnings) that were to be answered primarily in an open format. Because these questions are known to be highly sensitive and therefore usually yield comparatively high rates of item nonresponse, additional closed format questions with salary ranges were added. Thus, if respondents were not able or willing to report their salaries/earnings in detail, they were administered closed format questions. The corresponding list specified six salary ranges, which were defined by the average gross wage/salary in each country (for example, the third category for the German question on monthly salary was: 1 300 euros to below 2 300 euros). The German categories were determined using data provided by the Federal Statistical Office, based on the German Microcensus.

e) Countries/Languages
PIAAC included several questions relating to countries or languages, such as the country of birth or languages most often spoken at home. To facilitate the answer process for respondents, the most common countries or languages were given as a closed list, with an additional open format for countries or languages not included. These closed lists of countries and languages were adapted in each country. In Germany, the lists were adapted based on the most common countries of origin (citizenship) or languages, according to data provided by the Federal Statistical Office (Statistisches Bundesamt, 2008).

However, no adequate information on the country of birth or mother tongue was available. Data on citizenship were therefore used as a proxy to derive the categories for these questions.

f) Field of Education
In the international background questionnaire, the response option for questions on field of education was a closed list. Because the terminology used in this list was academically oriented, PIAAC Germany adapted this for respondents with a vocational qualification to a question with a free text entry. These respondents specified the name of their qualification instead of choosing a field of education.

g) Answer schemes related to occupational questions
The last set of adaptations in the German PIAAC questionnaire was related to modifications and extensions of certain response categories. For example, the categories of the question related to the type of employment contract held by the respondent (variable name: D_Q09DE) were adapted to include the full set of national employment contracts.

1 The examples were the same as those used in the SOEP 2008 questionnaire. A third question was included in the PIAAC field test, based on the German demographic standards. However, this was dropped for the main survey because most respondents did not provide any additional information.

After data collection, responses to the nationally adapted questions were (re-)coded into the international scheme (see Section 4.2).

In Germany, the adaptation process was supported by national background questionnaire experts and by the Scientific Advisory Board. The Scientific Advisory Board also participated in the selection process for the national extensions for Germany.

Questions about the following subjects were added to the German questionnaire as national extensions:

- part-time retirement,
- respondent's citizenship,
- place of residence before the fall of the Berlin Wall in 1989,
- language spoken at age 16,
- parental country of birth, and
- parental occupation.

The extension about part-time retirement was introduced to accommodate the German retirement system, in order to differentiate between respondents who were in the *active* versus the *passive* phase of their part-time retirement. Those in the active phase still work, whereas those in the passive phase do not. However, all persons in part-time retirement are officially regarded as *employed*. The wording of the questions in all sections related to employment was also adapted to accommodate this situation.

The other extensions were chosen for national analytical purposes. For example, the place of residence before the fall of the Berlin Wall is crucial for analyzing differences between respondents who grew up in the former German Democratic Republic versus the Federal Republic of Germany. Parental occupation was included in order to obtain important information on aspects of the respondents' socio-economic background and social mobility.[8]

The German background questionnaire can be downloaded from the GESIS PIAAC website (http://www.gesis.org/piaac/).[9] In the background questionnaire documentation, questions that were adapted are identified by an additional country code included in the variable name (in Germany: DE). Extensions are identified by the country code and an additional X (in Germany: DEX). The German adaptations and extensions are not included in the international *Public Use File*, but they are available in the German *Scientific Use File* (see Section 4.3).

2.2.2 Testing the Background Questionnaire

Following the completion of the translation and adaptation process, and once again after the changes made between the field test and main survey, countries tested the background questionnaire. In Germany, thorough testing was carried out, focusing on: (a) routing, (b) computation of derived variables, (c) overall flow of background questionnaire, (d) export files, (e) CAPI functionality and navigation, (f) stability of CAPI system, and (g) review of final translation.

Two approaches were followed to test the routing:

1. Each background questionnaire question was tested individually, i.e., each possible response was entered and the subsequent corresponding routing checked against the CAPI specifications.[10] Due to the routing complexity, numerous paths through the background questionnaire had to be tested.
2. All these paths were compared to a national background questionnaire HTML[11] file, in which the routings were implemented as hyperlinks, as well as to a graphical representation of the background questionnaire flow.

All derived variables created by the CAPI system by combining information from different variables were tested to check whether the allocation process was correct. This was especially important in Germany, due to the number and the complexity of the national derived variables for the education questions.

The overall flow was tested using a number of pre-defined biographies. Some of these scenarios were internationally designed and adapted for the German background questionnaire, and additional scenarios were developed and tested for national purposes. Furthermore, mock interviews were carried out. The flow tests verified whether the correct questions were administered, and whether the content and numbering of show cards were consistent. Functionality tests involved, amongst other things, checking navigation through the background questionnaire (using the keyboard or mouse) and testing whether help functions worked. All corresponding export files were checked for completeness and correctness. Furthermore, the stability of the CAPI platform was tested. For this, testers tried

8 Originally, this question was assessed in the field test version of the background questionnaire in all countries. Due to time constraints, it was not included in the final version of the international background questionnaire.

9 It can also be found under doi:10.4232/1.11865 (or https://dbk.gesis.org/dbksearch/download.asp?db=D&id=52072).

10 An international tool enabling testers to jump to specific questions was used here.

11 HTML = hypertext markup language

to crash the system in various ways, such as using function or other keys indiscriminately, or clicking extremely quickly. Translation testing involved examining whether the wording used throughout the background questionnaire was harmonized (also through the different routings) and that each question was comprehensible in the overall flow. It also involved checking the so-called *dynamic text*. Dynamic text means that several questions were worded differently depending on the responses to previous questions. For example, when asked about their job, respondents who were not currently working received the question that used the past tense (thus referring to their last job) or, for respondents currently working, the question was worded in the present tense (thus referring to their current job).

All problems were reported to the international Consortium via a ticketing system. Bugs that were fixed had to be re-tested. This meant that, in several cases, numerous iterations were necessary.

2.2.3 Administering the Background Questionnaire

The background questionnaire was administered as a CAPI, so that interviewers were routed through the questionnaire. For reasons of standardization, interviewers were instructed to read the questions verbatim and not to allow respondents to look at the text on the laptop screen. Interviewers recorded respondents' answers on the laptop. Because some of the response lists were very long, interviewers had a set of show cards that they handed to respondents. Respondents could thus read the response categories on the show cards, which facilitated their response process.

For respondents with language problems, interviewers were encouraged to find an interpreter (e.g., a family member) to help translate the background questionnaire questions and responses. Note that while this was explicitly allowed for by the PIAAC standards for the background questionnaire administration, help of any kind—including translation support—was forbidden for the direct assessment.

2.3 Assessment Instruments

This section provides a brief overview of how the assessment instruments for literacy, numeracy, and PS-TRE were developed by the international Consortium, as well as some information on the German implementation. As mentioned earlier in this chapter, the PIAAC survey design foresaw that the assessment items were to be implemented, by default, on the computer. Implementation of the assessment on the computer has several advantages: It allows a wider range of constructs to be measured, and also yields rich paradata, such as mouse clicks and time stamps. Furthermore, measurement precision and efficiency is increased through adaptive testing. Respondents who did not fulfil the necessary requirements for the computer branch were administered the paper-based assessment.

2.3.1 PIAAC Skill Domains

The three basic skill domains measured in PIAAC—literacy, numeracy, and PS-TRE—represent key information-processing skills that are regarded as crucial in work-place, personal, educational, social, and civic contexts. International expert groups developed frameworks for each domain, defining the underlying constructs and their various dimensions. Based on these frameworks and in consultation with the expert groups, the international Consortium developed the assessment instruments for PIAAC. Analogous to the background questionnaire, the assessment items were developed in English and then translated and adapted by

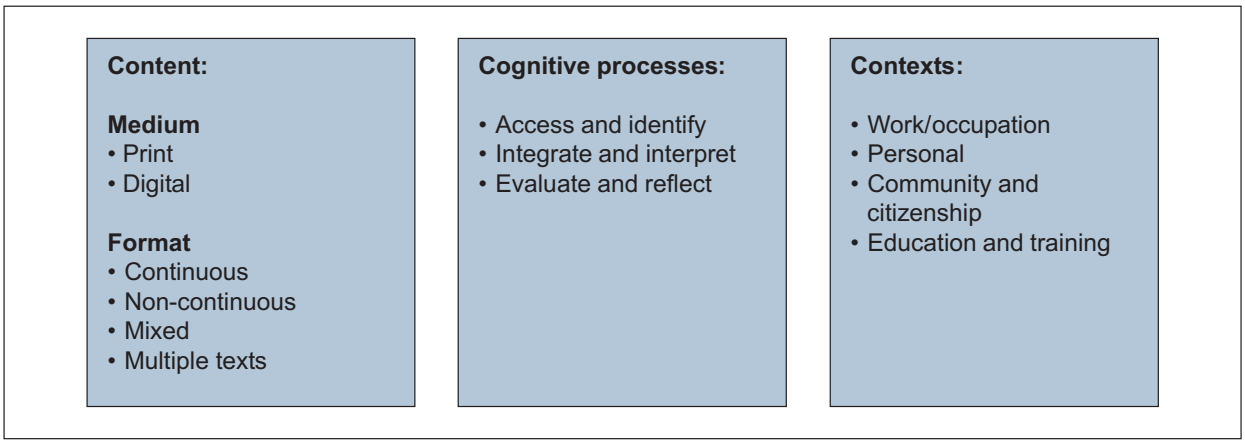
each country. This crucial element in instrument realization is described in detail in Section 2.4.4.

Literacy

Literacy is one of the key basic skills needed in today's society. As the amount of written information that individuals have to process in everyday life increases in quantity and complexity, demands on literacy skills are also growing. The international literacy expert group defined literacy for PIAAC as follows: "*Literacy is understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential.*" (Jones et al., 2009, p. 8).

The literacy framework for PIAAC (Jones et al., 2009) was based on the frameworks of the previous surveys on adult literacy, IALS and ALL. Whereas IALS differentiated between prose and document literacy, both ALL and PIAAC combined prose and document literacy into a unified construct. PIAAC extended the literacy concept to include literacy in the digital environment, for example, reading electronic text, which is becoming increasingly widespread in modern societies.

The literacy framework identified three central construct dimensions that define the content, the context, and the cognitive processes needed to manage literacy tasks successfully. These dimensions were the foundation for the literacy item development (a simplified scheme is shown in Figure 2.5).



Note. Based on Jones et al. (2009).

Figure 2.5. Dimensions of literacy

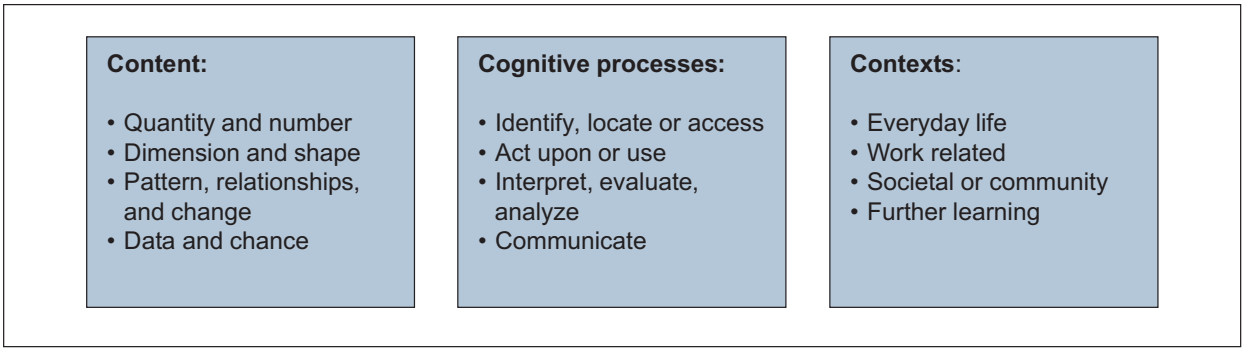
One of the goals of PIAAC was to provide a precise measure of literacy skills at the lower end of the literacy scale. Adults with low skill levels are at greatest risk for negative social, economic, and labor market outcomes. Previous surveys showed that a more detailed assessment at the lower levels of literacy proficiency had not been adequately addressed empirically (Jones et al., 2009). Thus, PIAAC introduced the reading components, which extended the literacy concept measured in PIAAC to include three additional basic types of tasks: *word meaning* (print vocabulary), *sentence processing*, and *basic passage comprehension* (Sabatini & Bruce, 2009). The reading components items were administered on paper only. They were an optional assessment component carried out in Germany and the majority of the participating countries.

Numeracy

Numeracy has long been considered a fundamental skill needed by adults to function effectively at work and in their private life. As technical and quantitative aspects of our everyday environment become more important, adults are increasingly required to cope with mathematical demands in a variety of situations.

In PIAAC, the definition of numeracy was widely based on the ALL numeracy framework, with a somewhat broader construct. The international numeracy expert group defined numeracy as follows: "***Numeracy** is the ability to access, use, interpret, and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life*" (Gal et al., 2009, p. 21).

The numeracy construct was also described along the dimensions content, cognitive processes, and contexts, as illustrated by the simplified scheme shown in Figure 2.6.



Note. Based on Gal et al. (2009).

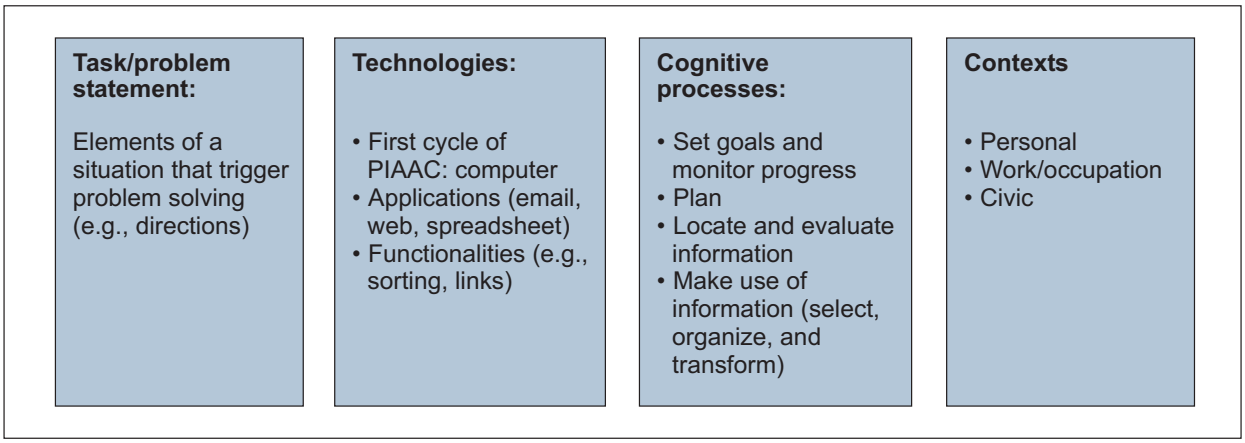
Figure 2.6. Dimensions of numeracy

Problem Solving in Technology-Rich Environments

The need for information-processing skills in the context of new technologies increasingly permeates modern life, and empirical information about this new set of skills is lacking. Thus, a new domain was developed to obtain information on how adults cope with everyday problems using new technologies. PIAAC is the first survey to assess the domain PS-TRE, which was made possible due to the computer-based implementation of the assessment. This international option was assessed in Germany and all but four of the other participating countries. As defined by the international expert group, PS-TRE involves:

> *using digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks. The first PIAAC problem solving survey will focus on the abilities to solve problems for personal, work and civic purposes by setting up appropriate goals and plans, accessing and making use of information through computers and computer networks.* (Rouet et al., 2009, p. 9)

It is important to note that, by definition, no assessment of PS-TRE was possible in the paper branch. Figure 2.7 shows a simplified overview of the dimensions that were differentiated for PS-TRE.



Note. Based on Rouet et al. (2009).

Figure 2.7. Dimensions of PS-TRE

For details about the competency domains, see the corresponding frameworks (Gal et al., 2009; Jones et al., 2009; Rouet et al., 2009; Sabatini & Bruce, 2009); for information on the item development and example items, see OECD (2013a, 2013b, 2013c); an overview of the domains and German example items for literacy, numeracy, and PS-TRE are given in Zabal et al. (2013). Examples of German reading components items for word meaning and sentence processing are shown in Annex A.2.2.

2.3.2 Item Development for PIAAC

The main objectives for the development of the cognitive instruments were to ensure that the items in the assessment (a) were well-balanced with respect to the construct dimension distributions recommended in the frameworks, (b) showed adequate psychometric properties, (c) were culturally appropriate and comparable across countries, and (d) yielded an adequate distribution across difficulty levels (see OECD, 2013b, 2013c). For more details on the development of the cognitive instruments see Lennon and Tamassia (2013).

The PIAAC cognitive instruments consisted of a number of items that were taken from previous surveys, as well as items that were newly developed for PIAAC. Approximately 60% of the PIAAC literacy and numeracy items were taken from IALS and ALL (Yamamoto, Khorramdel, & Von Davier, 2013a). These so-called linking items were included in order to link back to these previous surveys and enable trend analyses. These items remained generally unchanged for PIAAC. However, since IALS and ALL only used paper-based items, it was necessary to adapt them for their computer-based implementation in PIAAC. Various changes were required for the computer-based application; one major adaptation was the modification of the response modes. For example, for open text entry literacy items, a highlighting functionality, which could be automatically scored, was developed for the computer-based items (more information about automatic scoring is provided later in this section). Some of the items were implemented both paper-based as well as computer-based in PIAAC. The international Consortium analyzed data from the PIAAC field test to establish whether these two delivery modes were comparable. Results indicated no significant mode effects (OECD, 2013b); this was further confirmed by results from the main survey, which showed that the majority of items had common parameters for both the paper- and the computer-based versions of the same item (Yamamoto, Khorramdel, & Von Davier, 2013b).

In addition to the linking items, new items were developed by the international Consortium, in cooperation with the domain expert groups, and with contributions from countries.[12] At various stages of the development process, countries were asked to provide feedback on cultural and linguistic appropriateness and implementation aspects. All items for the new PS-TRE domain were developed specifically for PIAAC, as were the items for the reading components. It should be noted that the new literacy items were developed for use in one unique mode: either only computer-based or only paper-based. The new literacy items focused on electronic reading, as an innovative element of the literacy framework.

Table 2.1 shows the number of linking and new literacy and numeracy items delivered in one mode only, and the number of items that were administered in both modes. A complete list of linking units (consisting of one or more items) is provided in Annex A.2.3.

Table 2.1. Linking and New Items in Paper and Computer Mode

	Literacy		Numeracy		
Mode	Linking	New	Linking	New	Total
Only paper-based	/	6	2	2	10
Only computer-based	12	22	14	18	66
Paper-based and computer-based	18	/	17	3	38
Total	**30**	**28**	**33**	**23**	**114**

Note. / = no items.

As mentioned previously, one of the central aims of the field test was to test the functioning of the cognitive items. Therefore, the field test included more items than the main survey: (a) 85 items for literacy, (b) 75 items for numeracy, (c) 24 tasks in 14 scenarios for PS-TRE, and (d) 324 reading components items. The final main survey assessment instrument consisted of (a) 58 items for literacy, (b) 56 items for numeracy, (c) 14 PS-TRE tasks in 11 scenarios, and (d) 100 reading components items. Item selection was based on the international Consortium's analyses of the assessment items and the evaluation of their statistical performance, using field test data. Amongst other things, this involved checking whether items showed differential item functioning, i.e., were harder or easier in certain countries, compared to the rest. Based on these analyses, and with a view to balancing the item pool in terms of construct representation, the final item set was selected by the international Consortium and the domain experts for the main survey.

12 Germany did not submit any items.

2.3.3 Implementation of Computer-Based Items

Before working on the computer-based cognitive assessment, respondents received a short introduction (orientation) to the tasks, including information on functionalities and navigation options. Respondents were informed that they could not go back to previous items, once they had completed a unit, and were also told that some items were easier than others. For each specific domain, respondents learned about the different response modes.

The screens for the computer-based items had a consistent layout and display throughout all assessment domains: The left-hand side showed the task information, i.e., user instructions and *directions*, as well as the *questions*. The right-hand side of the screen displayed the *stimulus* material, i.e., the text, graphical information, etc., that needed to be processed to answer the question. *Navigation* and *help buttons* were located at the bottom of the left panel. Figure 2.8 shows this general layout (for a literacy or numeracy screen).

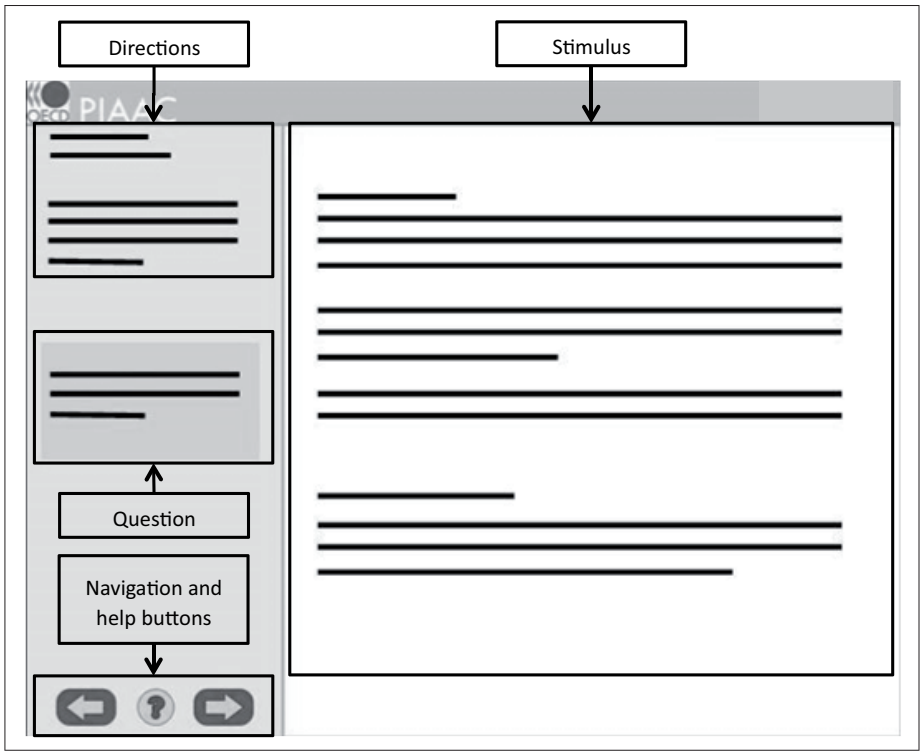


Figure 2.8. Example of a computer-based literacy or numeracy item layout

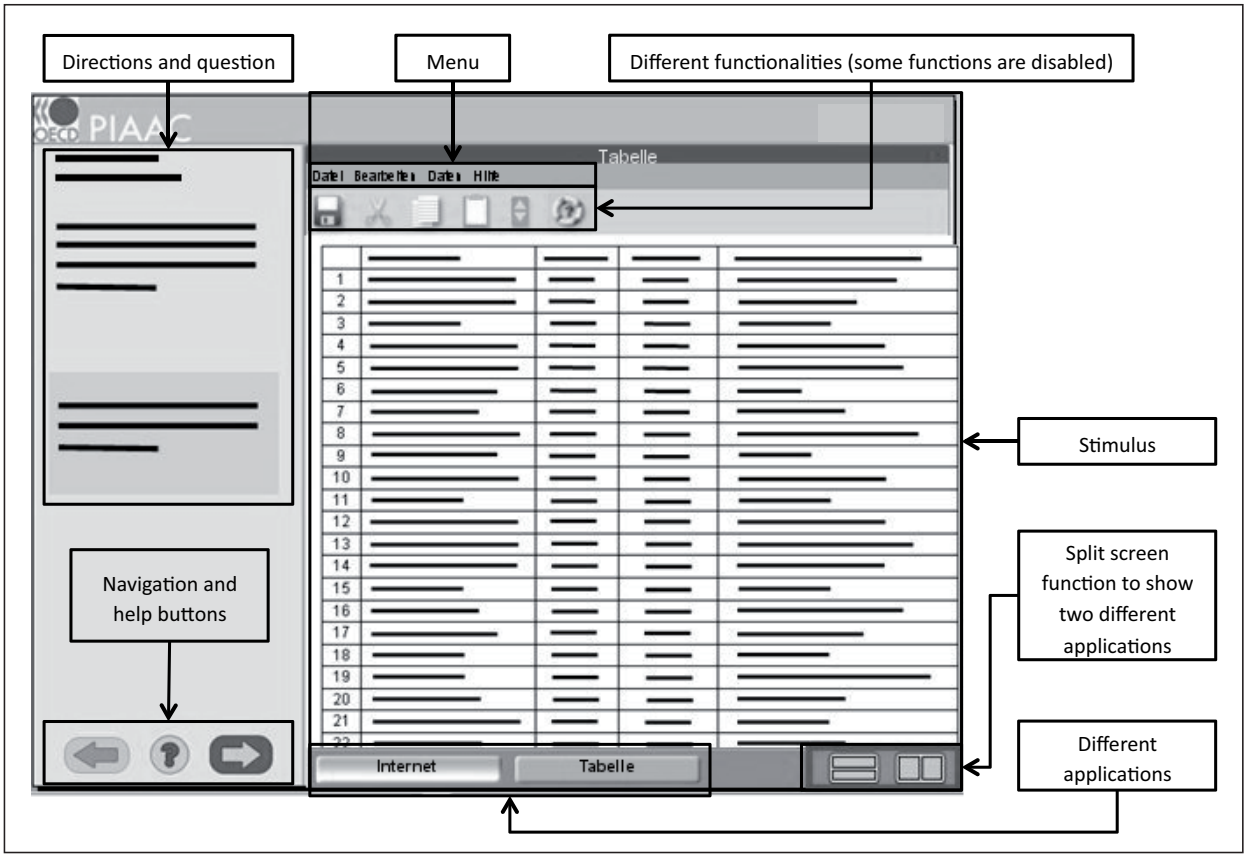


Figure 2.9. Example of a PS-TRE item layout

The screens for PS-TRE were more complex because the stimulus side also included various icons and navigation options; these were part of the technology tools (web, spreadsheet, email) and functionalities (e.g., the split screen function) made available for the PS-TRE scenarios. An example of the layout of a PS-TRE item with a spreadsheet environment is shown in Figure 2.9.

Sometimes the same stimulus was used for various items—this set of items corresponded to a unit.

For computer-based literacy and numeracy items, three main response modes, which generally required only rather basic computer skills, were available: clicking, numeric entry, and highlighting. A *clicking item* required the respondent to use the computer mouse to click on a certain part of the stimulus, or select multiple choice options or a radio button on the left panel. A typical clicking item, for example, involved clicking on a link or on a graphical element in the stimulus. A *numeric entry item* involved typing a numeric answer into the response box on the left panel, using the number keys, a period or comma for the decimal point, and the space key; all other keys were locked. Responses to *highlighting items* were given by using the computer mouse to highlight a part of the stimulus text in response to the question; required responses could range from highlighting one or more words to several sentences. In addition to these three response modes, a few items used a pull-down menu. Table 2.2 shows the distribution of items over the various response modes for literacy and numeracy.

Table 2.2. Response Modes of Literacy and Numeracy Computer-Based Items

Response mode	Literacy	Numeracy
Clicking item	16	16
Numeric text item	3	34
Highlighting item	33	0
Pull-down menu	0	2
Total	**52**	**52**

In PS-TRE, the response modes were more complex and included using pull-down menus, clicking on links, sending emails, and setting bookmarks.

As described in Section 2.1, the computer-based branch of the assessment delivered different testlets consisting of different sets of items. Testlets were always domain-specific, i.e., they contained items from one domain only.

Automatic Scoring of Computer-Based Items

In previous surveys, and also for the paper-based assessment in PIAAC, the respondents' answers to paper items were scored by human scorers after data collection. Scoring means that the response given by a respondent is evaluated as either correct or incorrect according to detailed scoring guides. The scoring definitions are an integral component of the items. For the computer-based assessment in PIAAC, however, in order to be able to implement an adaptive testing algorithm that achieves better estimates using a smaller number of assessment items per person, it was necessary to find a way to score the computer-based items via an immediate, automatic process. Concretely, it was necessary to evaluate the responses as correct or incorrect immediately after respondents had entered their answers. Because this task had to be carried out by the computer without interaction with a human scorer, exact rules for assessing whether an answer was correct or incorrect (the *scoring definitions*) had to be integrated into the programming of the computer-based items. Correct responses were scored as *1*, incorrect responses as *7*, and no response as *0*.

These scoring definitions were specified for the international literacy and numeracy items and had to be subsequently adapted in each country according to specific instructions (the process for PS-TRE was different and will be explained later). Scoring definitions were a part of the international verification process to ensure the comparability of the instruments, as described in Section 2.4 on translation.

The different response modes required different approaches to the scoring definitions. The *scoring definition for clicking items* specified in which area of the screen, for example on which link, the respondent had to click for a correct response. These items required only minimal adaptation, because click areas generally remained consistent across different national versions.

The *scoring definition for numeric entry items* could include a single number, a range, or a list of discrete numbers. In PIAAC, two different scoring methods were used: (1) number match: the numeric response was considered correct if it represented the same numerical value as the correct response, regardless of the way the number was "spelled" (1.5 = 1 ½), and (2) exact match, which meant that the scoring algorithm only accepted an exact answer (e.g., 1.0 was not accepted if 1 was considered the correct response). Adaptations to scoring rules for numeric entry items were required for items in which numbers or number formats had been nationally adapted. In Germany, the final instruments accepted different numerical spellings, for example, different decimal separators, to accommodate the increasing internationalization of number use.

For *highlighting*, scoring definitions included determining a minimum and maximum response area that specified the acceptable correct response. The minimum response area

was defined by the key information which was, per se, sufficient to qualify as a correct response. The maximum response area corresponded to the maximum information that could be highlighted and still remain an allowable correct response. This included the key information as well as additional, but non-contradictory, information. The rationale for this was that an otherwise correct response was disqualified when incorrect information was additionally highlighted. The response always had to include the minimum correct response (necessary condition). Using a special tool (the *Textblock Translation Editor*), two specific text blocks were technically defined: (a) the minimum correct response, and (b) the miss area that determined an incorrect response. The miss area was the complement of the maximum correct response. The highlighting response mode required the most adaptations, as it is very language-dependent. Due to the length and structure of the German language, finding comparable scoring definitions for some highlighting items proved to be challenging.

The scoring process for PS-TRE items differed from that for literacy and numeracy. Automatic scoring was not implemented for this domain because the items were not administered adaptively. Instead, the items were dichotomously and polytomously scored by the international Consortium, based on information in the output files (such as clicks on links, navigation through the website, etc.). Thus, scoring for PS-TRE took place after fieldwork, once the national main survey data were submitted to the international Consortium. At the same time, but prior to scaling, procedures to handle missing data for the literacy and numeracy items were applied by the international Consortium (Yamamoto, Khorramdel, & Von Davier, 2013c).

2.3.4 Implementation of Paper-Based Items

As described in Section 2.1, respondents who did not have sufficient computer experience or could not work on the computer-based items for other reasons were administered the assessment on paper. In total, there were four paper booklets: (1) core booklet, (2a) literacy booklet, (2b) numeracy booklet, and (3) reading components booklet. The core booklet contained eight comparatively simple questions (four literacy and four numeracy questions), whereas the literacy and numeracy booklets had 20 items each. The reading components booklet included three sections; each respondent had to answer a total of 100 short questions. Respondents were generally administered three booklets: (a) the core booklet, (b) either the literacy or the numeracy booklet (assigned at random), and (c) the reading components booklet. In some cases, respondents were administered only two booklets: Respondents who failed the core skipped the literacy or numeracy paper assessment and worked directly on the reading components.[13]

The literacy and numeracy paper booklets started with an introduction to the tasks. Respondents were instructed by the interviewer to read these before beginning to work on the tasks. In general, for most of the items, the stimulus was on the left-hand side of the booklet and the questions on the right-hand side. Analogous to the computer-based assessment, respondents answered each question without any time limitations. They were also encouraged to work through the booklet sequentially.

The paper items had several different response modes: open items, circle items, and multiple choice items. The literacy paper assessment (literacy core items and literacy booklet) included 22 open items, one circle item, and one multiple choice item. The numeracy paper assessment (numeracy core items and numeracy booklet) was composed of open items only.

As a part of the item development, the international Consortium produced item-by-item scoring guides specifying which responses were to be scored as correct and which as incor-

13 A small group of respondents who failed the CBA core were only administered the reading components booklet, as described in Section 2.1.

rect.[14] These scoring guides were subsequently nationally translated and adapted. After data collection, all paper booklets were scored by specially trained human scorers who followed the detailed PIAAC scoring instructions. The scoring process is described in more detail in Section 4.2.

For the routing in the paper branch, it was necessary to evaluate responses to core items immediately, because respondents who passed the core were to be administered the literacy or numeracy booklet, whereas those who failed only received the reading components. Therefore, it was necessary for the interviewers to score core booklet items during the interview itself. The interviewers were provided with the scoring rules on the laptop screen, and they had been extensively trained for this task during their training (see Section 3.2).

The reading components booklets consisted of three different types of tasks: word meaning (print vocabulary), sentence processing, and basic passage comprehension. For word meaning, respondents were shown a picture and four alternative words. Out of these, they were asked to circle the matching word. In sentence processing, respondents were shown simple sentences and asked whether they made sense. For basic passage comprehension, respondents had to read a short passage. Respondents were asked to choose one of two possible words to end some sentences within this passage, i.e., they had to circle the word that completed the sentence meaningfully.

Since all reading components items were multiple choice questions, a complex scoring process was not required. Responses to reading components were straightforwardly recoded as correct and incorrect after data collection.

2.3.5 Testing the Assessment

One important national responsibility was to test the general functionality of the computer-based assessment, to ensure that the national versions worked correctly. Testing had to be re-iterated several times during the development of the national instruments for the field test, because there were several deliveries of the instrumentation. Furthermore, problems that were reported and fixed had to be tested again. Another extensive round of testing was carried out after finalization of the instruments for the main survey.

In Germany, the computer-based assessment was tested in-depth. For some of these checks, testing scenarios were provided by the international Consortium; for others, we produced our own testing plans. Testing covered seven areas:

- *Checking the translation and readability of all texts on the screens*: This involved reading the text as a whole, comparing it with the original version of the item, checking whether the layout was correct and that all the text on the screen was visible, legible, and corresponded in font and layout to the international master.
- *Testing the routing across sections in the computer branch*: The different pathways through the assessment were systematically tested to determine whether the routing worked according to the design specifications.
- *Checking the correct allocation of testlets/adaptive testing*: The allocation of testlets depended on several entries in the background questionnaire, entries in the assessment, and a random factor. This test checked whether the adaptive testing design worked according to the assessment design specified by the international Consortium. Furthermore, the correct numbering of units and questions was checked.
- *Testing the functionality*: In the computer-based assessment and, in particular, in the PS-TRE items, many functions, such as using pull-down menus, search and sorting functions were available. The tests checked whether the functionalities worked correctly and whether the designated functionalities were available for each specific item.

14 Only one literacy item also had a partial credit.

- *Testing the automatic scoring*: Scoring testing was an elaborate and time-consuming process. The first tests of the scoring definitions were carried out on a testing portal conceived especially for this process. These tests were later repeated with the national virtual machine. To check that the technical implementation of the scoring definitions worked correctly, each item had to be tested numerous times by entering different correct or incorrect responses. For example, different spellings were tested for the numeric entry items. For highlighting items, the minimum and the maximum correct response were checked, as well as many other correct and incorrect answers.
- *Checking the export files*: This especially crucial check included ensuring that all corresponding files were correctly exported and, for example, that information on entered responses was accurately documented.
- *Testing the stability of the platform*: Similar to the testing for the background questionnaire, the stability of the software was tested.[15]

In comparison, checking the paper booklets was straightforward and consisted primarily of checking the correctness and legibility of the text, as well as the layout. A final version of the printed paper booklets was sent to the international Consortium for sign-off prior to final booklet printing. This ensured correct implementation and standardization across countries.[16] In addition, the functions and navigation of the international software for the paper branch had to be carefully tested, because the interviewers received instructions for the administration of the paper assessment in the international software.

2.3.6 Instructions to Interviewers and Material Given to Respondents

Interviewers were extensively prepared for their role as assessment administrators during their training. They were also strictly instructed *not* to help respondents during the assessment.

For the assessment, respondents were given a set of materials that they could use to answer the items. Specifically, they were given a numeracy kit consisting of a photo,[17] a calculator, and a ruler. Respondents were also provided with paper and a pencil, for notes. According to the international specifications, the calculator was only allowed to have simple functions. The calculator and the ruler both had to be signed-off by the international Consortium, once again to ensure that the same standardized conditions were met across all PIAAC countries. Photos were supplied by the international Consortium.

The interviewers were automatically routed through the different branches of the assessment and were given appropriate instructions by the international software. In the computer branch, interviewers were first instructed to read information about the assessment to the respondent. These instructions indicated that it was important that respondents attempt all tasks, but that some tasks were easier than others. Interviewers also announced that there would first be a short introduction to the assessment (the orientations) before the actual assessment started. Interviewers were then instructed to hand over the computer, the numeracy kit, and paper and pencil to the respondent. Respondents then worked on the computer-based items on their own and without any time restrictions.

In the paper branch, interviewers first read out a set of instructions. Respondents were told, for example, that if they could not answer a task, they could just move on to the next one (i.e., the answer to a particular question was not needed for the next question). The instructions in the international software guided the interviewers, who first handed

15 As a part of the instrument testing, international and national scripts controlling the virtual machine were also extensively tested.
16 As part of the standardization across countries, even the weight of the paper to be used for the booklets was specified by the international Consortium.
17 The photo was required for a specific numeracy item.

respondents the core paper booklet, paper, and a pencil. The interviewers had to score the core items, and the international software directed them through this process and provided appropriate scoring instructions. If the respondent had completed the core booklet successfully, the interviewer was instructed to hand over the literacy or numeracy booklet—the international software randomly selected one of the two booklets, i.e., the corresponding booklet was automatically displayed in the instructions to the interviewer. If the numeracy booklet was assigned, the interviewer was instructed to give the respondent the numeracy kit. Once the respondent had completed the (literacy or numeracy) booklet, the interviewer was instructed to hand over the reading components booklet and to read out the appropriate instructions. Because the rate of processing information was of particular interest in the reading components assessment, interviewers were instructed to click on a timer button on the computer screen as the respondents started each task and to end timing when they completed it.

2.4 Translation of Instruments

Adequate translation and adaptation of survey instruments are essential to ensure the equivalence of measurement and, thus, comparability across different languages and cultures. PIAAC recognized the need to establish best-practice procedures and quality control measures in order to produce translated instruments of the highest possible quality in all the PIAAC languages. Towards this aim, the international Consortium produced a set of translation and adaptation guidelines and recommendations, sophisticated translation tools, and in-person training (see Ferrari, Wayrynen, Behr, & Zabal, 2013).

All master versions of instruments and materials were in English.[18] Countries were responsible for the translation into their national language(s). In Germany, the survey instruments were translated only into German. The country versions were subsequently verified by the international Consortium, i.e., their quality was vetted by independent linguistic experts trained in the evaluation of instrument translation.

Three different sets of materials with somewhat different requirements were translated:

- the background questionnaire,
- the direct assessment, and
- survey administration materials.

As recommended in the international guidelines, a double translation and reconciliation procedure was implemented in Germany both for the background questionnaire and the assessment material. Translations were produced by professional translators. A committee or teamwork approach was used for reconciliation of the various versions, and the final steps and reviews were carried out by the German National Center. The challenge of reconciliation consists in merging the two independent translations into a final version that achieves the best possible balance between faithfulness and fluency (Ferrari et al., 2013). Reconciliation using a teamwork approach has the advantage that the mix of skills and expertise needed for creating the best possible translations can be brought together.

There were no particular specifications for the translation of the survey administration materials (specifically, the very comprehensive interviewer manual and interviewer training materials). These were first translated by a professional translator and then extensively re-worked, with diverse national amendments and supplements, by the German National Center, in collaboration with the survey organization.

18 Only one source version was produced for PIAAC (unlike PISA, which had source versions both in English and French).

The full translation process described in this section took place in preparation for the PIAAC field test in 2010.[19] Following the analyses of the field test data by the international Consortium, only minimal changes to the German translations were required and implemented for the main survey.

Before describing the translation of the background questionnaire and the assessment in more detail, we will first outline the selection of translators, their training, the translation tools, and the translation workflow in PIAAC.

2.4.1 Translators

All translators involved in the translation of the PIAAC instruments for Germany were required to have the following qualifications:

- skilled practitioners, with perfect command of source and target languages and cultures,
- German (target language) as mother tongue,
- experience in translating questionnaires (for translation of background questionnaire),
- experience in the translation of cognitive instruments and, thus, some familiarity with the principles of assessment (for translation of assessment), and
- an adequate level of computer literacy.

The first four national requirements were mandatory. The last was desirable but not obligatory. In total, four professional translators were commissioned for the translation of the PIAAC instruments in Germany. Two of them produced separate translations for the background questionnaire, and two translated the direct assessment.

The international Consortium provided in-person translation training to assure the international comparability of the translation process. This training familiarized translators with survey objectives and basic design features, the PIAAC translation workflow, and item-by-item specifics. For Germany, the two professional translators responsible for the translation of the assessment, the translation coordinator, and various other members of the National Center attended this training. Furthermore, the German National Center and the national IT coordinating team gave the translators of the assessment a supplementary one-day training session. This national translation training consisted of hands-on technical training and a review of central translation recommendations and procedures. Translators were provided with the full set of documentation relevant for their work.[20]

2.4.2 Translation Material, Workflow, and Tools

The international Consortium produced general translation guidelines as well as item-by-item guides for both the background questionnaire and the direct assessment. Furthermore, the translation process was supported by the international Consortium via a ticketing system that could be used by countries to consult international reviewers and item developers. The translation process was managed by the countries via a so-called Item Management Por-

19 As described in the previous sections, more background questionnaire questions and assessment units and items were tested in the field test than were actually used in the main survey, because one of the objectives of the field test was to test the functioning of the PIAAC measurement instruments in all countries and to select the best set of questions and items for the main survey (see also Annex A.1).

20 The two translators entrusted with the translation of the background questionnaire did not receive the PIAAC translation training as provided by the international Consortium, but they were given detailed instructions. The schedule of the German translation of the background questionnaire preceded the international schedule, because several members of the German National Center were also responsible for the translation guidelines for the background questionnaire as a part of their involvement in the international Consortium.

tal. This was a secure online facility (web portal) available to all PIAAC user groups involved in the translation process: translators, national centers, verifiers, and other members of the international Consortium. The Item Management Portal gave users access to all translation-relevant documents (from guidelines to electronic item files) and managed all necessary file exchanges between the users. This web portal also offered a preview function, which was of utmost importance for the assessment material.

The national translation workflow encompassed the following basic phases and feedback loops:

1) initiation of translation by German National Center (installing translator accounts and allocating translation assignments to individual translators, thus giving them access to electronic files),
2) translation; release of translations by translators to German National Center,
3) team reconciliation and review of translation by German National Center; release of finalized translation to verifier,
4) verification (international quality control); release of verifier feedback to German National Center,
5) various feedback loops between German National Center, verifier, and other members of the international Consortium (including several layout checks and subsequent corrections; for assessment material, also various processing loops for scoring implementation), and
6) final check.

The initial translation phase (first three steps listed above) took approximately three to four months. Approximately five additional months were required after the first verification (steps four to six). The entire process was documented in several spreadsheets.[21] These contained the English source version, sub-divided into the stimulus, the interviewer instructions, questions, and scoring definitions. They also specified the translation recommendations (so-called item-by-item directives), verifier interventions (verifiers allocated severity codes when they identified problems with national translations, and also proposed corrective actions), further country comments (including requests for national adaptations), and Consortium responses for the subsequent verification loops, up to the final check. Sometimes the target version was also inserted in English, to facilitate communication with the international Consortium and test developers, because these could obviously not be familiar with all survey languages. Thus, back translation was used here only for documentation purposes.

The electronic files for the computer-based material (background questionnaire and computer-based assessment) could be processed using Computer-Aided Translation tools such as TRADOS or the Open Language Tool (OLT). Translations were therefore produced in XLIFF files with an XML interchange format.[22] In Germany, the OLT XLIFF translation editor was used for translation processing. In this editor, the text segments on the left-hand side showed the source version, and the translations were inserted into the identical segments on the right-hand side. Translation of the paper-based material was done with Word files.

21 Background Questionnaire Adaptation Spreadsheet and, for each direct assessment unit, an individual Verification Follow-up-Form

22 XML = Extensible Markup Language; XLIFF = XML Localization Interchange File Format

2.4.3 Translation of the Background Questionnaire

The content of the background questionnaire required a certain amount of structural adaptation to obtain both nationally appropriate and internationally comparable background information. Structural adaptations included, for example, changes in the number of response categories or even, to a certain extent, modifications of the questionnaire routing. Given the technical constraints of the translation process (i.e., using XLIFF-files as the basis for translation), structural adaptations had to be finalized before the provision of the translation files. Information regarding the German structural adaptations to the background questionnaire is provided in Section 2.2.1. Here, we will focus only on the actual translation; this may include non-structural adaptations to a certain (relatively minor) extent.

The general aim in translating the background questionnaire was to transport the same semantic content, question intent and design features into the target language in a way that was (equally) comprehensible to a very heterogeneous group of respondents. This also included catering to regional variations in language or differences in vocabulary use in different age groups. Beyond the usual translation challenges, the terminological consistency of key terms and recurrent wording was extremely important. Therefore, during the entire translation process, key terms, instructions or response scales used repeatedly in different sections of the questionnaire received special attention. Furthermore, some systematic differences in wording were implemented via use of dynamic text. For example, the wording of questions relating to parental occupation differed, for 16 year olds, from that for the rest of the respondents.[23] Dynamic text was entered into a special spreadsheet[24] and then used to create the national electronic XLIFF files for translation.

If the international background questionnaire adopted questions from an existing international survey in which Germany had participated, we took the existing, validated German survey questions (see Section 2.2). The questions were then critically checked, because adherence to the PIAAC international source version was a prerequisite that had precedence over comparability with existing surveys. When necessary, changes were made so that the German PIAAC translation complied with the PIAAC source version.

As mentioned above, two independent translations were produced by professional German translators experienced in translating survey questionnaires. The background questionnaire was reconciled in a team review consisting of: the two professional translators; two national background questionnaire experts; several in-house GESIS experts (with expertise in survey translation, questionnaire design, and survey methodology); members of the German National Center, including the national translation coordinator, who led the reconciliation process. The reconciliation meeting reviewed the two translations for each question (including those that were taken from other surveys), all response categories, and interviewer instructions. A decision was made either for one of the two translations, a combination thereof, or a new version was created by uniting the best features of the individual translations.

For the background questionnaire, the follow-up revisions and finalization process were protracted. The entire questionnaire translation was carefully reviewed to check the quality of the final translation and adherence to the PIAAC background questionnaire translation guides. Another review focused on harmonization between all the background questionnaire sections. Furthermore, it was important to ensure that the terminology and question wording of the structural adaptations were consistent with the rest of the questionnaire and that the overall flow was smooth. The national background questionnaire experts were reconsulted about major changes.

---

23 For the 16 year olds, the introduction read: "Does your mother or female guardian hold a paying job?", whereas, for the others, it read: "Did your mother or female guardian hold a paying job when you were 16 years old?".

24 Dynamic Text Rule Spreadsheet

2.4.4 Translation of the Direct Assessment

For the direct assessment, the aim of translation and adaptation was to achieve cognitive equivalence. Translating the assessment requires a different focus and skills set than the translation of questionnaires, because it requires a basic understanding of the principles of test construction. For example, it is critical that translations avoid changing the item difficulty by making them simpler (e.g., by providing unintentional clues) or more difficult. It is important to recognize critical item characteristics and to be aware of translation-relevant differences between domains. For example, the items for numeracy explicitly used relatively simple language to reduce the overlap with the literacy domain. At the same time, it was essential to understand whether mathematical concepts were expressed in technical or in non-technical language, because this needed to be reflected in the translation. For PS-TRE, the language used was to be more informal and colloquial to fit the item context.

A general issue arose from the fact that, due to characteristic features of the German language, the German translations were typically longer than the English source texts. This was especially problematic for the computer-based assessment: It was crucial that all language versions be comparable in terms of the screen set-up. For example, it was important that German respondents should not be obliged to scroll more often, or at different locations, than in the source version and other PIAAC languages. Thus, this was an additional challenge that made it necessary to re-work translations in later phases, if the degrees of freedom allowed for in the international layout were not sufficient to accommodate the length of the German text.[25]

A few adaptations were internationally prescribed, the most notable of which were required for the numeracy assessment. The national versions were to display numerical information as was customary in the country, and use appropriate units of measurement and the national currency. In Germany, for example, we used commas as decimal separators, implemented the metric versions of items, and changed the currency signs to euros.[26]

The translation of the reading components was somewhat different from that of the other domains. Due to the unique linguistic characteristics of each language, the translation of the reading components required significantly more adaptations. For example, it was often necessary to construct language-specific distractors. Specific instructions and adaptation procedures were provided in the translation guidelines for the reading components.

For the translation process, it was important to differentiate between the new items that were developed for PIAAC and the linking items from IALS and ALL. We will first describe the translation process for the new items.

For the new parts of the direct assessment, the following sets of materials were translated:

a) new literacy items (including paper scoring guides),
b) new numeracy items (including paper scoring guides),
c) the full set of PS-TRE items,
d) the full set of reading components (including paper coding guides),
e) orientation modules,
f) assessment workflow, which was especially extensive for the paper branch because it also included all the scoring instructions for the interviewers, and
g) ICT core.

For the translation of the direct assessment (sets a–d), two independent translations were produced. For the assessment material, the German National Center shared translations with

25 In a few cases, some deviations were unavoidable and could not be solved through translation changes.
26 There was no need to change the monetary amounts, since the dollar (source) and euro were sufficiently comparable.

the Austrian National Center.[27] Thus, a German translator produced one translation, and the Austrian translation served as a second independent translation in the double translation plus reconciliation procedure used.

As mentioned above, for each domain, the two translations were reconciled in a team review with the following participants: the German translator, one national expert per domain, an in-house GESIS expert on survey translation and members of the German National Center.[28] For the reading components, linguistic expertise was especially important, and an in-house GESIS linguist therefore also attended the reconciliation meeting.

The materials in the sets e–g (mentioned above) were first centrally translated by the international Consortium, and then carefully reviewed and re-worked by the German National Center.

For the IALS and ALL linking items, the procedure was different, because extensive direct translation work was not involved. Because Germany had participated in IALS, original versions of the German IALS items were available and were carefully reviewed by the German National Center. These items were essentially left unchanged with the following exceptions:

a) If deviations from the international source or the international PIAAC guidelines were found, or if errors in the original IALS-DE version needed to be corrected, the international Consortium and the literacy expert group were consulted about these changes, decisions were reached bilaterally, and the approved corresponding changes finalized.
b) Some adjustments were required by the recent German spelling reform.
c) For the computer-based items, some further modifications were internationally required and implemented, according to the specifications in the corresponding PIAAC guidelines.

For the ALL linking items, the procedure was somewhat different, because Germany did not participate in this survey. The national experts first reviewed the Swiss-German versions of the ALL items (Notter, Arnold, Von Erlach, & Hertig, 2006).[29] Based on these reviews, the German National Center revised and implemented the expert recommendations, and then reviewed all items. As an additional source of information, the modified ALL items used by the Institute for Employment Research (IAB) in their ALWA study (Working and Learning in a Changing World; Kleinert, Matthes, & Jacob, 2008) were consulted, as were translation experts for final touches. Otherwise, as for the IALS items, the international Consortium and expert groups were consulted when clarifications were needed, compliance to the PIAAC guidelines was enforced, and the spelling reform was implemented.

As was the case for the background questionnaire, the German National Center extensively reviewed and re-worked all translations after the reconciliation meeting. At this stage, both the new assessment materials as well as the linking items were carefully processed. This process covered:

- reviewing all translations (including the paper scoring guides, if applicable) and their adherence to item-by-item guides. This included checking the text and displaying all computer-based items via a preview function, to ensure that the stimulus, questions, and instructions were realized as intended.

27 Sharing translations between the two countries greatly reduced the costs. However, it was only feasible for the assessment because there were too many national systemic differences for this to work for the background questionnaire.

28 The Austrian translator was not available for the reconciliation meeting. For literacy, the domain expert was only available by phone.

29 Due to the fact that work on the linking items began before GESIS was appointed to carry out the National Project Management for PIAAC in Germany, this review was commissioned by an interim National Project Manager.

- harmonizing recurring terms over all sets of materials (e.g., consistency across units in directions, help menus, navigation, etc.).
- checking version consistency for items that were delivered in both the computer- and paper-based modes.
- defining automatic scoring definitions for the computer-based literacy and numeracy items. For example, for items with a highlighting response format, this was implemented in the finalized national XLIFF file using the Textblock Translation Editor (more information on automatic scoring is given in Section 2.3.3). The automatic scoring definitions were also included in the international verification.
- processing of verification results, implementation of corrections and numerous parallel rounds of layout checks (as mentioned previously, layout issues sometimes needed to be addressed by finding a shorter translation). Sometimes, such changes also entailed changes to the automatic scoring. This phase included several feedback loops with the international Consortium.

2.4.5 Changes to Translations Prior to Main Survey

After the field test, some revisions to the measurement instruments were required by the international Consortium. Additional changes could also be requested by the countries. This process was documented in the relevant main study spreadsheets.[30] Results of the international field test item analyses carried out by the international Consortium were also documented here, i.e., national items with poor item functioning were identified and countries asked to check for possible translation errors. The German National Center accordingly scrutinized the translations of the few items that were flagged by the international Consortium, and made changes if a significantly improved version was found. Overall, only a limited number of changes were required and made to the field test instruments. These subsequently underwent a verification by the international Consortium.

2.5 Technical Platform and Specifications

Although CAPI is standardly used in many surveys, PIAAC was the first international large-scale survey to also implement the assessment on computers. This innovation raised various technical challenges that were successfully addressed by the international Consortium together with the PIAAC countries, so that the main survey was run on a stable, well-functioning platform.

At a national level, it was necessary to bring together different areas of expertise, to establish the technical groundwork and to ensure a smooth interview process, as well as the production of reliable and error-free data outputs. In Germany, the technical implementation work for PIAAC was shared between three organizations. The national IT coordination was handled by the DIPF. It had the overall responsibility for national IT processes and provided second-level technical support for the survey organization. TNS Infratest, the survey organization, was responsible for the case management system, integration of the international software into the case management system, and other company-specific systems and software. It also provided the hardware and delivered first-level technical support for interviewers. The German National Center at GESIS was responsible for the overall coordination of technical issues and for testing the software components.

30 The Main Study Translation-Adaptation-Verification Monitoring Form, a spreadsheet similar to that used for the field test.

The following sections present an overview of the international software infrastructure for PIAAC, the German technical workflow, different national hardware and software components, and specifics of the German case management system.

2.5.1 International Technical Components

The PIAAC technical environment was relatively complex, because it incorporated both the background questionnaire and assessment components into a virtual machine, which was then integrated into the external national case management system. The decision to implement the PIAAC survey instrument as a virtual machine was made to ensure the comparability of the PIAAC delivery, irrespective of computer operating systems, because the virtual machines are well-separated from the surrounding host systems. All software embedded into the platform for the international PIAAC survey instrument consisted of open source components. The operating system within the virtual machine itself was a Debian Linux system. On this Linux platform, several software components were combined to realize the survey interview, which consisted of the background questionnaire and the assessment module. The main part of the software embedded into the virtual machine was the open source assessment platform TAO (*Testing Assisté par Ordinateur*),[31] which was the outer shell for test delivery. The assessment items in the TAO delivery system itself were based on different technologies, such as the survey module developed especially for PIAAC, *TAO Qual*, for the background questionnaire, the *CBA-Itembuilder* for literacy and numeracy items, and the *TAO BLACK* model for the PS-TRE items.

First, a so-called *Mother VM*[32] was built that contained all common data, such as the operating system, and all country-independent software components. Second, the base national virtual machines were derived from this Mother VM, and country-dependent information and data were retrieved from the Item Management Portal (see Section 2.4.2). These components were then assembled into the individual national virtual machines and delivered to the countries. They were then tested in several rounds until they reached a final delivery status (for more detailed information, see Upsing et al., 2013b).

2.5.2 German Technical Workflow

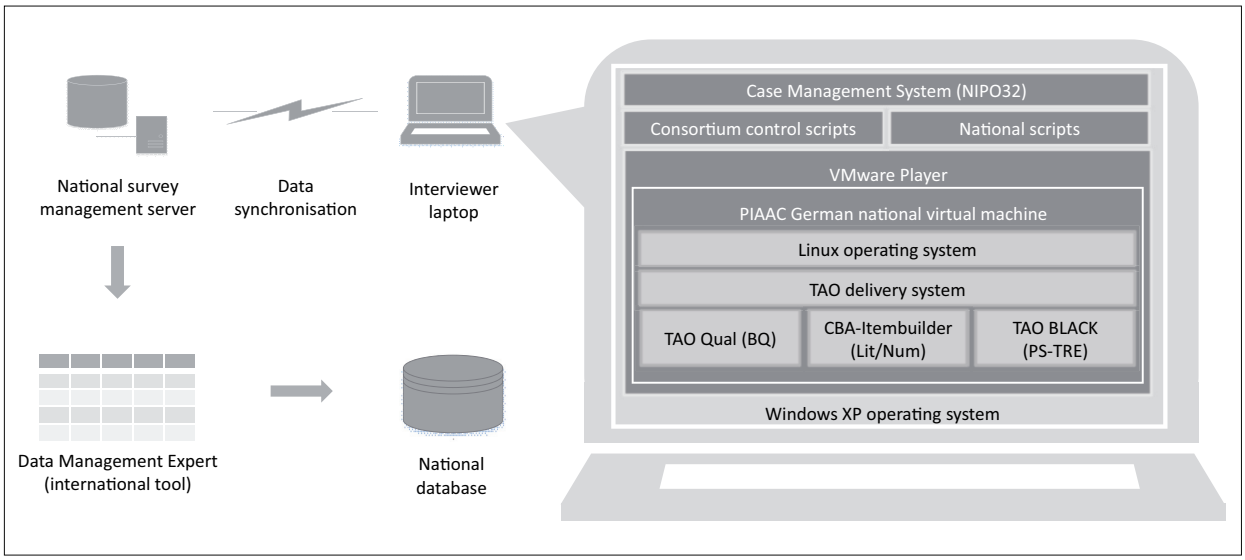
The PIAAC software infrastructure in Germany consisted of three main components:

- PIAAC virtual machine,
- case management system NIPO32, and
- Data Management Expert software.

Figure 2.10 illustrates the general PIAAC technical components and the software installed on the interviewer laptops. The survey organization home office transferred case information and assignments from the National Survey Management Server to the case management system on the interviewer laptops. All interviewer laptops were equipped with a Windows XP operating system, the case management system, the VMware Player, the German virtual machine, and both international and national scripts. Interviewers regularly transmitted interview export files for each finalized case back to the National Survey Management Server. Subsequently, the Data Management Expert software was used to handle the interview export files (see Section 4.1 for details about the data processing steps).

31 The current version of TAO can be downloaded from http://www.taotesting.com/resources/download-tao.

32 VM = virtual machine



Notes. BQ = background questionnaire. Lit = literacy. Num = numeracy. TAO = *Testing Assisté par Ordinateur.*

Figure 2.10. Technical PIAAC components and German interviewer laptop software

More concretely, in the interview situation, interviewers started a case via the case management system client on the laptop. After this case initialization, the case management system launched a script that subsequently started the virtual machine and imported the required case parameters, such as the respondent's age and gender, into the virtual machine. After completion of the PIAAC interview, the case management system triggered the virtual machine to produce export files and to shut down.

The technical workflow, including all components, was tested exhaustively prior to the fieldwork period.

2.5.3 Virtual Machine and Scripts

In Germany, the national virtual machine was run on the VMware Player. As mentioned above, the virtual machine was well-separated from the surrounding system. However, some restricted communication between the host operating system (Windows XP) and the guest (virtual machine) was necessary, for example, to (re-)start or stop the virtual machine and to exchange data. Therefore, the international Consortium provided a number of scripts to control the virtual machine. These scripts interacted with the PIAAC virtual machine via the VMware VIX interface software. This software provided services such as controlling virtual machines, file handling, and calling programs and scripts inside the virtual machine. The German technical implementation made use of all the international PIAAC scripts for virtual machine remote control, case management, and data access, as well as maintenance and administrational requests. A list of these scripts can be found in Annex A.2.4; a detailed description of the PIAAC scripts and their functionalities is summarized in the international technical report (Upsing et al., 2013b).

The virtual machines also incorporated a patch mechanism which could be used to fix software bugs in the virtual machine after installation on the interviewer laptops. Contrary to the field test, the German virtual machine proved to be stable in the main survey and patching was not required during fieldwork.

A national feature of the German virtual machine was a splash screen that was displayed while the virtual machine was shutting down. This splash screen covered a basic case management system code and reduced the risk of unwanted interviewer interactions

with the case management system (thus reducing possible system destabilizations and crashes) until the shutdown and data export were completed.

2.5.4 Case Management System

The case management system used by the national survey agency TNS Infratest is called NIPO32 (also known as COMPASS or ODIN) and was provided by TNS NIPO (Netherlands) and TNS Infratest (Germany) to carry out the PIAAC survey in Germany and in the Netherlands. The case management system was used by home office to assign cases and supervise interviewers, and by interviewers to manage their assignments on a day-to-day basis. Nevertheless, NIPO32 was not built to handle the complexity of an embedded virtual machine. Furthermore, as mentioned above, the international instrument did not allow, for security reasons, any direct communication between the external software on the laptop and the TAO system embedded into the virtual machine. Therefore, it was necessary to develop another software layer between the software components. This third specific software component was a national script, MOAS (*Mother Of All Scripts*), which was used to control relevant scripts from the international set of scripts.

2.5.5 Interviewer Laptops

The PIAAC standards and guidelines required that the interviewer laptops used for the interview meet specific hardware requirements (OECD, 2010b). The standard for the screen size was especially crucial, because the computer-based assessment items needed to be comparable across countries, and deviations in display and stimulus size could lead to psychometric differences impacting on the assessment results. The standards for the laptops were defined by the international Consortium at a very early stage of PIAAC (beginning of 2008), and strove to define requirements that would also be feasible for countries with a relatively old IT infrastructure (i.e., purchased well before the start of the study). Meeting the international standards proved to be a challenge in many countries, as was the case in Germany. The laptop characteristics defined in the standards were no longer available when the laptops were purchased for the field test data collection in Germany. The main issue was that the screen size of 1024x768 pixels at 14” specified by the standards could not be met, because all vendors had changed to a widescreen format in the interim. Therefore, the international Consortium adjusted the screen size standards to account for these changes and allowed limited degrees of freedom. To ensure that the standardization was met across countries, national laptops had to be approved by the international Consortium prior to the field test fieldwork.

In Germany, *Fujitsu ESPRIMO Mobile D9510* laptops were purchased with a screen size of 1280x800 15.4”. This provided 94% of the size of the original 1024x768 14” screen and was within the allowed 10% deviation from the original screen size. The German laptops had the following specifications:

- Fujitsu ESPRIMO Mobile D9510,
- 15.4 inch WXGA Notebook (native resolution 1280x800 pixel),
- Intel Core 2 Duo P8700, 2.53 GHz,
- 2GB RAM,
- 160GB HDD,
- Intel GMA X4500HD,
- DVD+- DL RW,

- Windows XP Professional Version 2002 with Service Pack 3 installed, and
- Fujitsu Eco Black Wired Optical Mouse.

The German interviewer laptops complied fully with the international standards; the hardware specifications actually exceeded the recommended parameters indicated in the international standards.

In addition to these standards, the settings in Germany were harmonized on all laptops to ensure that the same conditions applied. This included (a) deactivating the standby mode, (b) setting the screen brightness to 100%, (c) disabling the scroll wheel of the computer mouse, and (d) deactivating the touchpad.[33]

33 The touchpad was deactivated to avoid unintentional cursor movements and highlighting during the assessment.

3. Sampling, Fieldwork, Weighting, and Nonresponse Bias Analyses

PIAAC strove to implement best practices in survey methodology in all participating countries. At a national level, the German National Center undertook serious efforts to find the nationally optimal survey design and procedures for the different stages of the survey life cycle, so as to minimize survey errors from different sources. This included defining and implementing the best sampling design to realize a probability-based sample as well as fieldwork procedures and strategies for maximizing response rates while monitoring and limiting nonresponse, all of which are of special importance in high-quality face-to-face surveys.

At an international level, the quality assurance process required countries to document their survey design and processes in numerous forms. For example, very early in the process, an extensive National Survey Design and Planning Report, encompassing critical aspects of the standards and guidelines, had to be submitted by participating countries.

This core chapter of our national technical report elaborates on the German sample design and selection (Section 3.1), fieldwork strategy and survey operations (Sections 3.2 and 3.3), weighting of the data and subsequent nonresponse analyses to evaluate the quality of the national data (Section 3.4).

3.1 Sample Design and Selection

PIAAC was designed to provide insights into the skills and competencies of working-age adult populations, specifically adults aged 16 to 65 years. Each country participating in PIAAC was required to specify a sample design that generated a probability-based sample that was representative of this target population (OECD, 2010b). The international Consortium implemented a set of quality control checks and procedures that were performed throughout the entire sample selection process, in order to assess and evaluate all sampling activities in the participating countries (Mohadjer, Krenzke, & Van de Kerckhove, 2013b). For this purpose, detailed sampling design plans and several forms documenting the sample selection process were prepared by each country; these forms needed a final sign-off by the international Consortium.

In Germany, a registry-based, two-stage stratified and clustered sampling design was implemented, with the selection of (1) municipalities as primary sampling units (PSUs) or clusters at the first stage and (2) individuals at the second stage. Given this design, each in-scope unit had a chance of being selected for the sample and, thus, for each sampled unit (PSU or person), a non-zero probability of selection can be calculated.

The present section provides general information on the target population, sampling frame, sample size, sample design, and sample selection, as well as information specific for Germany.

3.1.1 Target Population, Sampling Frame, and Sample Size

The PIAAC *target population* consisted of non-institutionalized[1] adults aged between 16 and 65, whose usual place of residency was in the country, here Germany, during the time of data collection, regardless of their citizenship, legal status, or first language (Mohadjer et al., 2013b; OECD, 2010b). For all countries with registry samples (including Germany), all adults aged 16 to 65 years on 1 December 2011 (reference date) were included.

The *sampling frame* is a list or another type of register from which a sample is drawn. In a multi-stage sample design, a sampling frame is required for each stage of selection. In Germany, the sampling frame for the first selection stage was the Federal Statistical Office frame of communities, dated 30 December 2009.[2] For the second stage, the registers of the local population registries in the municipalities that had been selected at the first stage were used as sampling frames.

Given that not all individuals of the target population in Germany are properly registered[3] at their local population registry (e.g., illegal immigrants), the lists of inhabitants used did not completely cover the target population. According to international standards, sampling frames were to cover at least 95% of the target population (Mohadjer et al., 2013b; OECD, 2010b). For Germany, this standard was met: The noncoverage rate for illegal immigrants was estimated to be about 0.5%,[4] supplemented by another 2% for other individuals who were not listed in the register, for example, because they had recently moved and not yet registered at the local registry of their new place of residence.[5]

For the Consortium to be able to calibrate reliable item parameters and to define a population model for each test language in PIAAC separately, a minimum *sample size* of 5 000 completed cases[6] was required (this refers to countries that included the optional domain PS-TRE, as was the case for Germany). Germany started with a gross sample of 10 240 selected individuals. The achieved net sample comprised 5 465 completed cases.

3.1.2 Sample Design and Sample Selection in Germany

According to PIAAC standards, a self-weighting sample design had to be implemented, in order to obtain a sample that was representative of its target population. Countries had to adhere to widely recognized principles of scientific sampling "to achieve the maximum precision possible for a given sample size, while limiting the costs of data collection." (OECD, 2010b, p. 44). No substitution of sampling units was permitted at any stage (OECD, 2010b).

In Germany, the methodically best option available was selected: a registry-based sample design, which had previously been successfully implemented in various high-quality national social science surveys, e.g., the German General Social Survey, ALLBUS (see Koch, 1997; Wasmer, Scholz, Blohm, Walter, & Jutz, 2012).

---

1 Adults residing permanently in institutions, such as prisons, nursing homes, or military bases, were excluded from the target population.

2 The Federal Statistical Office frames of communities (*Gemeindeverzeichnisse der Bundesrepublik Deutschland des Statistischen Bundesamtes*) for different years can be obtained from the DESTATIS website www.destatis.de.

3 Proper registration refers to the principal residence and not to a secondary residence.

4 This is a rather conservative estimate that was based on information on the total stocks of irregular foreign residents in Germany retrieved from the Database on Irregular Migration (Hamburgisches Weltwirtschaftsinstitut, 2010).

5 The 2% refer to ex post exclusions allowed for registry samples. The information about individuals not listed in the register was obtained during data collection as a part of the address search procedure to trace respondents (see Section 3.3).

6 For the definition of a completed case, see Section 3.3.

First Stage: Selection of Municipalities

For the first stage, a stratified random sample of municipalities was selected with a *probability proportionate to size* (PPS) design. First, municipalities were stratified to cells of a matrix that was formed by a combination of the following variables: federal states, administrative regions, districts, and urban/rural status (using a variable called BIK).[7] Sample points were subsequently allocated to the stratification table. Sample points encompass a fixed number of individual addresses to be drawn from the registries of the selected municipalities at the second stage.

Measure of size (MOS) for all municipalities was the overall number of individuals in the target population. The allocation of sample points to the stratification cells occurred in proportion to the MOS of the strata. The MOS for each stratum was derived from the total MOS of all municipalities in the relevant stratum. Sample points were allocated to strata by applying Cox's controlled rounding algorithm (1987). This procedure ensured that the distribution of selected municipalities in the sample reflected the distribution in the whole population with regard to federal states, administrative regions, districts, and rural/urban status.

After application of the Cox algorithm, the selection of a municipality within a sampled stratum was achieved by systematic random sampling with a random start and a predefined sampling interval. In general, there was one sample point per municipality. However, more than one sample point was allocated to larger municipalities, due to their MOS.[8] Overall, 320 sample points were selected in 277 municipalities.

Second Stage: Selection of Persons

After stage one, registries in the sampled municipalities[9] were contacted and informed about the PIAAC survey. In general, municipalities in Germany are not obliged to select persons and their addresses for the purpose of scientific research. In the past, in other surveys based on registry samples, municipalities that refused to provide addresses were substituted by municipalities with similar characteristics from the same sampled stratum (Wasmer et al., 2012). As mentioned above, substitution in PIAAC was not permitted at any stage. Thus, in Germany, methods had to be established to convince representatives of reluctant municipalities of the necessity to participate. This was achieved by several approaches, both conducted by the survey organization or the German National Center, and ranged from phone calls to personal visits by team members at local registries. Overall, there was only a small number of reluctant municipalities. In the end, the efforts were successful and all municipalities agreed to provide addresses.

The selected municipalities were provided with the following instructions for the selection process:

- Draw a systematic random sample of individuals, with a random start number and a sampling interval.
- Select 180 cases per sample point in municipalities with 500 000 or more inhabitants, or 120 cases in municipalities with 100 000 to 499 999 inhabitants; otherwise, select 60 cases.
- Provide personal information (such as name, address, age, gender, nationality) for each selected case.

---

7 The information was derived from official statistics of the Federal Statistical Office from 30 December 2009.

8 More than one sample point was allocated in Berlin, Bremen, Cologne, Dortmund, Dresden, Düsseldorf, Duisburg, Essen, Frankfurt/Main, Hamburg, Hanover, Leipzig, Mannheim, Munich, Nuremberg, Stuttgart, and Wuppertal.

9 Due to very high fees for an address selection in Braunschweig, Frankfurt/Main, Hanau, and Offenbach, addresses were not selected from the complete register in these municipalities, but from randomly selected city districts.

Data delivered by registries were checked for (a) consistency (e.g., selection only of valid individuals belonging to the target group), (b) completeness, and (c) duplicate copies. Inconsistencies were reported back to the registries and occasionally new data were re-delivered.

Following this cleaning process, the survey organization drew a stratified sample of individuals from the complete pool of delivered addresses and selected 32 cases per sample point. Selection was performed as follows: All delivered addresses for each sample point were first allocated to a 12-cell-matrix (six age groups by gender). Subsequently, an optimization algorithm was run for each sample point. This algorithm selected 32 persons under the constraints that (a) the socio-demographic structure of a municipality was well reflected with regard to age group and gender and (b) the socio-demographic structure of the federal state in which the municipality was located was also well matched. The remaining addresses were kept for use as reserve samples; ultimately, no reserve sample needed to be released.

Unfortunately, an error occurred at this stage. Usually, the survey organization's standard procedure is to randomly order the pool of addresses before drawing the sample. However, the cases had been sorted by age for quality control checks prior to selection, and this sorting order was mistakenly still in place when the sample was selected. In combination with a rounding error in the selection program, some age groups (those ending with 0) were over-represented, and others (in particular, those ending with 9) were under-represented in the sample. This discrepancy was detected only after data collection had been completed and thus could not be resolved by drawing a new sample. As a consequence, the sample no longer has the characteristic of an EPSEM[10] sample, as originally intended. Instead, the selection probabilities for the individuals vary.

Although it is, in principle, possible to model the incorrect selection probabilities, this process is enormously complex and time-consuming. Therefore, the selection probabilities were approximated through simulations, i.e., the erroneous optimization algorithm was repeated 10 000 times. In order to estimate the selection probability of each element of the sample frame, the frequency with which an element occurred in the 10 000 samples was counted. With this simulation approach, the resulting probabilities are adequate approximations of the "true" inclusion probabilities.

Thus, the sample remains probability-based (Mohadjer et al., 2013b; Mohadjer, Krenzke, Van de Kerckhove, & Hsu, 2013c) and is representative of its target population. It is a random sample with non-zero and known probabilities for all elements. The overall inclusion probabilities were determined as the product of the (known and positive) probabilities of selecting a municipality multiplied by the inclusion probabilities for the individuals in the selected municipality (as determined by 10 000 simulation runs).

3.2 Preparing for Fieldwork

In Germany, a variety of best practices were established for fieldwork and put in place to produce high quality, comparable data, and to achieve the best possible adherence to the international standard requirements, within national constraints. One of the most ambitious standards, from a national point of view, was the fieldwork standard for the required response rate. The minimal overall response rate for PIAAC was targeted at 70% (OECD, 2010b). Response rates between 50% and 70% were also regarded as acceptable if countries provided analyses indicating that the potential nonresponse and undercoverage biases were within acceptable limits. These benchmarks clearly exceed response rates realized by other high-quality international and national probability surveys in Germany. For example, the European Social Survey (ESS) achieved response rates of around 31% and 34% in the last

10 EPSEM = Equal Probability of Selection Method

two rounds in Germany (European Social Survey, 2012, 2013). Analogously, the German General Social Survey (ALLBUS), which is one of the leading social surveys in Germany in terms of survey methodology, obtained response rates of approximately 40% and 34% in the last two rounds (Wasmer, Scholz, & Blohm, 2010; Wasmer et al., 2012). Therefore, maximizing response rates—while simultaneously minimizing nonresponse bias—was one of the major focuses of the German fieldwork strategy.

The national fieldwork specifications for PIAAC were developed cooperatively during extensive discussions between the German National Center, GESIS fieldwork experts, and the national survey organization. The comprehensive set of fieldwork measures for PIAAC united national best practice, PIAAC standards and recommendations, and some supplementary provisions to achieve the best possible fieldwork results. The German PIAAC fieldwork strategy focused on:

a) *interviewers,* by
   - appointing experienced interviewers with excellent track records,
   - providing extensive interviewer training and equipping them well for the interview, and
   - offering attractive interviewer remuneration.

b) *addressing respondents,* by
   - providing appealing incentives for respondents and
   - developing engaging accompanying study materials (e.g., an informative brochure and flyer).

c) *enhancing the fieldwork processes,* by
   - implementing an elaborate contact strategy,
   - optimizing national fieldwork design and operations,
   - thoroughly supervising and monitoring fieldwork, and
   - implementing a comprehensive validation strategy, including 100% verification of all completed interviews.

We will now describe the German PIAAC fieldwork specifications with respect to interviewers and addressing the respondents. Details on how the fieldwork organization and processes were optimized will be given in Section 3.3.

3.2.1 Interviewers

Interviewers play a key role in surveys, both as recruiters of survey participants and as data collectors. Obtaining respondent cooperation requires special contacting and communication skills, ranging from locating the address, through showing perseverance in contacting and answering questions about the survey professionally, to creating an atmosphere of trust and interest necessary to gain respondent cooperation. Once the cooperation has been secured, interviewers need to conduct the interview following standardized procedures, while remaining respectfully attentive to individual needs and flexible enough to adapt to unexpected occurrences during the course of the interview. In PIAAC, the interviewers also had to switch smoothly from an active question-asking role for the background questionnaire to a more passive-observational role during the assessment. For the latter, interviewers were required to create a quiet and supportive atmosphere that allowed the respondents to work through the assessment at their own pace and without feeling pressured by the interviewer's presence. At the same time, interviewers had to remain watchful, in case

they needed to intervene and come to the aid of the respondent, for example, if technical problems occurred.

To optimize interviewer work and minimize interviewer effects, special care was taken in selecting and training interviewers. Given the fact that interviewers needed to be equipped with new interviewer laptops, to meet the technical standards for PIAAC (see Section 2.5), and due to other operational considerations such as workload, training, and availability, it was decided to employ 130 interviewers for the PIAAC survey. TNS Infratest chose the best 130[11] interviewers experienced in conducting high-quality social surveys with registry samples from its interviewer staff (consisting of more than 1 200 CAPI interviewers). Selection criteria were:

- excellent track record, especially for contacting and performance (response rates),
- proximity to PIAAC sample points, to ensure efficient and intensive work on assignments,
- availability during fieldwork period and for trainings,
- consideration of workload and concurring assignments, and
- seniority and previous experience with PIAAC field tests.

Table 3.1 shows the distribution of gender, age, and years of experience (years working as interviewers for TNS Infratest) of the German PIAAC interviewers.

Table 3.1. PIAAC Interviewer Characteristics

Interviewer gender	**%**
Male	53.5
Female	46.5
Interviewer age	
Up to 49 years of age	17.1
50 to 59 years of age	27.9
60 to 69 years of age	46.5
70 years and over	8.5
Experience as TNS interviewer	
Up to 3 years	24.0
4 to 10 years	47.3
11 to 20 years	13.2
21 years or more	15.5
Total interviewers	**100.0**

Note. Numbers were provided by TNS Infratest.

The large majority of German PIAAC interviewers were over 50 years of age and had been working for TNS Infratest as interviewers for a considerable period of time.

Interviewer Training

One important way to improve the quality of interviewers' work is through training. The international Consortium developed a concept and comprehensive materials for a five-day in-person interviewer training. While such extensive interviewer trainings seem to be com-

11 One interviewer was unable to work for PIAAC at short notice; thus, fieldwork for the main survey was actually carried out by 129 interviewers.

mon practice in North America, there is no tradition of lengthy interviewer trainings in Germany.[12] The very positive feedback from the five-day interviewer trainings for the PIAAC field test showed, however, that it is not only possible to carry out these trainings in Germany, but also that they are essential to the implementation of a complex survey, such as PIAAC, and were very well received by the interviewers and fieldwork supervisory personnel.

All selected interviewers participated in interviewer training for the main survey, as required by the PIAAC standards. Prior to training, interviewers were sent a Home Study Exercise, which introduced the interviewers to PIAAC, outlined key design and procedural elements of PIAAC, and also provided more general information about survey implementation. Interviewers were asked to read the document (which was approximately 50 pages in length) and complete the exercises at the end of each chapter, in preparation for the in-person training. The Home Study Exercise consisted of approximately 15 hours of at-home training. Completed Home Study Exercises were collected and checked at the interviewer trainings, with very good results.

Because only experienced interviewers worked on the PIAAC assignments, the interviewer training did not include comprehensive generic training, such as general interviewing techniques,[13] but focused on PIAAC-specific issues and standardized procedures.

Interviewers who had already been involved in the PIAAC field test received a somewhat condensed three-day version of the PIAAC interviewer training (22 hours of in-person training attended by 38 interviewers). The remaining interviewers without PIAAC field test experience, but expertise in face-to-face social surveys, received the full five-day training (31 hours of in-person training attended by 91 interviewers). Three identical interviewer training sessions were held in a conference hotel facility near Munich during the last two weeks of July 2011 (just before fieldwork started) and in the second week of August 2011 (these interviewers started fieldwork in mid-August, directly after their training).

The PIAAC interviewer training in Germany addressed the following topics: [14]

- general introduction to PIAAC (1.5 hours),
- overview of fieldwork phases and schedule (1 hour),
- gaining respondent co-operation (3 hours),
- contacting and documenting contacts, interviews, and non-interviews: the PIAAC contact protocol and the PIAAC disposition codes (2 hours),
- case management system: PIAAC-specifics (1.5 hours),
- CAPI training and technical tricks and tips (0.75 hours),
- the PIAAC background questionnaire (5.75 hours),
- introduction to the PIAAC assessment (1 hour),
- the computer-based assessment (1.5 hours),
- the paper-based assessment, including scoring the literacy and numeracy paper core (4.75 hours),
- practice interviews (6.25 hours),
- administrative procedures and quality control (1.25 hours), and
- summary and wrap-up (0.75 hours).

12 The German Social Survey did not conduct any interviewer trainings whatsoever in the last two rounds (Wasmer et al., 2010; Wasmer et al., 2012). The ESS in Germany carried out half- to one-day interviewer trainings for round 6, but no interviewer trainings for round 5 (European Social Survey, 2012, 2013). The National Educational Panel Survey has carried out one-day and two-day interviewer trainings (e.g., Aust et al., 2012; Aust, Gilberg, Hess, Kleudgen, & Steinwede, 2011).

13 However, a review of general interviewing techniques was included in the Home Study Exercise.

14 The number of hours allocated per session (indicated in the brackets) refers to the full interviewer training agenda for interviewers without previous field test experience.

Training materials developed by the international Consortium were translated, adapted, and extended for national purposes. There were 1–2 lead trainers and 3–4 assistant trainers per session, with one technical support person present at all times. Group sizes were between 18 and 31 participants. All lead trainers and the majority of the assistant trainers received training directly from the international Consortium, so that it was not necessary to hold an additional train-the-trainer session. Lead trainers from TNS Infratest (project directors) convened a special supervisory training session for their home-office staff. Additionally, PIAAC supervisors and other home-office staff took turns attending the in-person interviewer trainings.

Lead trainers from TNS Infratest were responsible for all the sessions pertaining to fieldwork administration procedures, contacting, and quality control. Lead trainers from the German National Center were responsible for all units pertaining to the administration of the PIAAC interview, i.e., background questionnaire and assessment. The training sessions included many hands-on exercises. For example, because the PIAAC interview was implemented with a special technical platform, it was necessary to familiarize interviewers with the new software. All sessions focused on how to use the international software, including launching, pausing, and closing the software, as well as using the various keys, functions, and navigation options. Furthermore, although interviewers were familiar with documenting contact information, the PIAAC contact protocol, called the case folder, and the disposition codes differed in many respects from the usual contact protocols. Therefore, various scenarios were developed and interviewers were required to fill out case folders accordingly.

The sessions on the background questionnaire included: (a) a review of standards in asking survey questions and recording responses accurately and completely, focusing on the importance of following study protocols; (b) an overview of the PIAAC questionnaire sections, walking the interviewers through main sections of the questionnaire in an interactive fashion, discussing additional interviewer instructions, and probing as appropriate; and (c) additional in-depth instructions for the questions related to education (one of the key variables in PIAAC), as well as eliciting adequate responses to the open-ended occupation and industry questions.

The sessions on the administration of the assessment provided an overview of the assessment design, explained the components of the assessment in detail, taught the interviewers how to score the literacy and numeracy paper core, and briefed interviewers on their special role during the assessment. Here, too, practical exercises and interactive units were used. The implementation of the paper assessment required special attention, because it involved many more administrative tasks for the interviewer than the essentially standalone computer-based assessment. At the end of training, the administration of the PIAAC interview was practiced in two role-play exercises, with each interviewer acting once as the respondent, and once as the interviewer. These role plays were essential for bringing together and consolidating all the PIAAC elements learned in the previous sessions.

The sessions on gaining respondent cooperation were set up so that interviewers could share their at-the-doorstep experience. Where possible, this session was offered jointly for PIAAC field test interviewers and interviewers new to PIAAC. Topics included coping with general contacting problems as well as with more PIAAC-specific problems, such as those arising from the assessment component (e.g., target persons' fear of being "tested"). In a concerted brainstorming session, the interviewers produced a set of possible reasons to participate in PIAAC. Additional role plays helped develop a variety of possible strategies for gaining respondent cooperation and deflecting respondents' fears or concerns. Interviewers were able to learn from each other and expand their contacting strategies, not only for PIAAC, but also for other surveys.

In addition to the regular training program, optional supplemental sessions were offered in the evenings, in which interviewers were coached according to their individual needs.

The trainer team also held daily review and debriefing sessions to discuss training progress and the interviewers' performance.

At the end of the training, interviewers were asked to fill out an evaluation questionnaire. The evaluations for all training sessions were very positive. The trainings not only provided an excellent preparation for the complex PIAAC interview, but also created an opportunity for interviewers to get to know each other and exchange notes on their fieldwork experience, and also for project staff and interviewers to bond. As a consequence, one of the additional bonuses of the interviewer trainings was that they created a unique PIAAC team feeling, motivating all participants to do their best for this challenging project.

Because there was no significant interviewer attrition during the eight months of fieldwork, additional training sessions were not necessary, although plans for such trainings were in place, in case of need. In a few cases, specific re-trainings or refreshers were carried out by TNS Infratest home-office staff or senior interviewers on a one-to-one basis. New issues that arose during fieldwork were communicated in writing to all interviewers, i.e., during fieldwork, interviewers received regular memoranda with updates and specific instructions from home office.

Interviewer Manual

Interviewers were provided with two important accompanying documents: an interviewer manual and an interviewer booklet. The interviewer manual was based on a document produced by the international Consortium that was translated and then re-worked and adapted quite extensively. The German PIAAC manual was over 180 pages long and included the following sections:

- *introduction and overview*: general information on PIAAC (e.g., objectives), international setting, national organization structure, timeline, sample specifications, interviewer responsibilities, ethical aspects, and confidentiality,
- *preparing for fieldwork*: overview of PIAAC materials,
- *interview workflow*: interview situation, starting the case management system, specifics of the background questionnaire, administering the computer-based assessment, administering the paper-based assessment (including interviewer scoring of the paper core), closing/pausing/breaking off interviews,
- *quality control*: tape recordings, monitoring, interview validation,
- *documentation in case folder and case management system*: contact rules, detailed explanation of the case folder, recording contact attempts, overview of disposition codes, assigning final disposition codes, and examples of completed case folders (these illustrated various scenarios for contact documentation and disposition code assignments),
- *finalizing and mailing paper case documentation back to home office*: checklist, post-processing specifications, and
- *annexes*: additional information on background questionnaire, suggestions for gaining respondent cooperation (including a list of frequently asked questions and suggested answers), information on interviewing respondents with impairments or special needs, tape recorder instructions, and laptop information.

All interviewers were required to read the manual. During fieldwork, it was an important reference document with which interviewers could check up on many aspects and specifics of the PIAAC interview.

For their day-to-day work, the interviewers were provided with an additional short reference document (interviewer booklet).[15] This 10-page document provided a hands-on, well-visualized, and succinct description of important information. Furthermore, it included a

15 This was a national document only and was not based on an international master.

short form of the scoring instructions for the paper core. Although scoring was trained intensively during interviewer training, it remained one of the most difficult (but nevertheless crucial) interviewer tasks. The interviewer booklet also included a practical overview of the PIAAC CAPI functions and keys, as well as a number of technical instructions and recommendations. It gave step-by-step instructions on how to close the interview, and how to pause or breakoff an interview. Furthermore, directions for using a tape recorder to record two full interviews (for quality assurance) were provided (for more information, see Section 3.3). Finally, it included a checklist of the complete set of materials needed for a PIAAC interview.

Interviewer Remuneration

Interviewer payment is an important determinant of interviewer motivation. The PIAAC standards and guidelines specified that interviewer remuneration for PIAAC was not to be based on a rate per completed interview, but rather per hour. Furthermore, it was to be attractive compared to other studies and related to the length and complexity of the interview (OECD, 2010b).

However, it is necessary to make allowance for the fact that survey organizations usually have standard policies for interviewer payment arrangements and it is rarely feasible to deviate significantly from these for a specific survey (see Stoop, Billiet, Koch, & Fitzgerald, 2010). The basic interviewer payment scheme at TNS Infratest is per complete interview; this also applies to face-to-face interviews. Several factors contribute to this arrangement: interviewers work as freelancers, this form of payment is transparent for both parties, fieldwork costs can be more reliably estimated, and the payment scheme is attuned to the standard validation strategy, which is based on the validation of completed interviews.

Therefore, careful thought went into determining the structure and an adequate level of interviewer payment for PIAAC that would:

a) be regarded as attractive by PIAAC interviewers, reflect the length and complexity of the interview, and stress the prominence of the survey;
b) compensate for the generally low survey cooperation in urban areas (and thus compensate for additional interviewer burden, due to the increased contacting efforts required for such sample points);
c) include an hourly payment component, specifically to ensure that respondents would be allowed the time they needed to complete the assessment;
d) allow a certain flexibility in payment for non-standard assignments; and
e) remain within the budget of the survey.

As a result, a mixed payment scheme was developed that featured:

- an attractive rate per completed interview,
- an add-on of approximately one sixth of the rate per completed interview for assignments in municipalities with 100 000 – 499 999 inhabitants,
- an add-on of approximately one third of the rate per completed interview for assignments in municipalities with 500 000 or more inhabitants,
- an additional per hour component for interviews that were longer than 105 minutes in total,
- reimbursement of all travel costs,[16] and
- day rates (for exceptional situations).

In summary, compared to other social surveys in Germany, the interviewer remuneration for PIAAC was generally higher and, while based on a per-piece scheme, it also included both extra payments for large municipalities, which are generally more work-intensive for

16 Reimbursing travel expenses is common practice for face-to-face surveys conducted by TNS Infratest.

interviewers, as well as a dynamic component related to the length of the interview. Thus, the spirit of the PIAAC standards was adequately achieved within the national German constraints.

3.2.2 Addressing Respondents

In addition to deploying excellent interviewers, other widely used measures, including incentives, advance letters, and study materials for respondents, can further enhance survey participation.

Incentives

Respondents received an incentive of 50 euros in cash upon completion of the interview; this amount remained constant in all phases of the data collection. It was decided to pay a 50 euro conditional incentive for the main survey after analysis of the field test data, in which PIAAC Germany tested three different conditional incentives: 50 euros, 25 euros, and a 10-euro commemorative coin. Fifty euros is a very attractive incentive in Germany, noticeably higher than what is usually offered.[17] Feedback from interviewers indicated that the conditional incentive was, indeed, very useful in obtaining participation.

An unconditional incentive was also distributed: Sticky notes with the PIAAC logo were enclosed with the advance letter. The rationale was that these would be useful for all types of individuals and households, promote familiarity with the PIAAC survey, and thus trigger recognition and indirectly promote interest in PIAAC when the interviewer made in-person contact.

Advance Letter

The advance letter is the first point of contact with the target person. For PIAAC, the advance letter was personalized: It was adressed to the target person by name and announced the upcoming contact by an interviewer, also introduced by name. It outlined the aim of PIAAC and its international context, explained how the target persons were selected and why their participation was crucial, and prominently specified the 50 euro incentive. At the same time, the letter stressed that participation was voluntary, and explicitely confirmed that data privacy laws would be adhered to. The advance letter introduced GESIS as the National Project Management and TNS Infratest as the survey organization. Annex A.3.1 shows the standard version of the advance letter in German. The advance letter was sent by mail to all target persons; attached were a flyer, a separate confidentiality/data privacy statement, and the sticky notes with the PIAAC logo, mentioned above, as an unconditional incentive. As a result, the envelope with the PIAAC advance letter and accompanying information set itself off from regular promotional mail, with the explicit intention of increasing the probability that target persons would actually open it and read the documents.

In order for the respondents' receipt of the advance letter to coincide as closely as possible with the interviewers' visits, these letters were sent at various pre-defined dates that had been previously synchronized with interviewers' contacting plans. Implementing a staggered mailing schedule was non-standard for the survey organization.

---

17 In comparison, between the years 2002 and 2010, the German General Social Survey offered either no incentive or a 10-euro commemorative coin, or 10 euros in cash (Blohm, Harkness, Klein, & Scholz, 2003; Blohm & Koch, 2013; Wasmer et al., 2012). The German ESS offered incentives in rounds 5 and 6: In 2010, respondents received 20 euros (European Social Survey, 2012), whereas, in 2012, participants were initially offered 25 euros; this amount was later increased to 40 euros, due to fieldwork difficulties (European Social Survey, 2013). The German ALWA study in 2007/2008 (Working and Learning in a Changing World), which included an assessment component, used a 15 euro incentive (Antoni et al., 2010). In 2010/2011 participants of the National Educational Panel Study (starting cohort 6) received 25 euros in cash (Aust et al., 2012).

Information Material

For PIAAC in Germany, we produced an appealing and professionally designed flyer and brochure that introduced the survey and encouraged the target persons to participate. The flyer summarized some central facts about PIAAC, including why the results were important, how the target person was selected, and why participation was crucial. The time period in which data would be collected was indicated and the PIAAC interview (e.g., average interview duration, types of questions and tasks) was described briefly. Finally, the flyer informed the readers that they would receive 50 euros, as a token of appreciation. This flyer was a part of the advance information mail sent to the target person, in preparation for an interviewer's visit. It is shown in Annex A.3.2 and can be downloaded under www.waxmann.com/buch3113.

The more elaborate brochure provided more details about the survey. It was essentially an at-the-door instrument and was quite widely used by the interviewers; the brochure can be found in Annex A.3.3 and downloaded under www.waxmann.com/buch3113. This more detailed brochure included:

- a *preface* by the National Project Manager and the President of GESIS, addressing target persons and encouraging participation in PIAAC;
- *general information about PIAAC*, i.e., PIAAC's aims, participating countries, importance of an international comparison, role of the OECD, national funders, similarities and differences between PIAAC and the *Programme for International Student Assessment* (PISA), expected date of publication of results; and
- *answers to frequently asked questions about participation*, i.e., who can participate and how were persons selected, adherence to data privacy and confidentiality, length of PIAAC interview, types of questions and tasks, reasons for participating in PIAAC (mentioning the 50 euro incentive), and links and phone numbers under which further information could be obtained.

Respondents who were interested in more information than that offered in the flyer and brochure could access the GESIS PIAAC website or call a hotline about their individual questions. Additionally, a press release was distributed to specific newspapers at PIAAC sample points, to increase the survey's public visibility and credibility for respondents.

In summary, interviewers had a very comprehensive set of at-the-doorstep materials to support them in gaining cooperation:[18] the PIAAC flyer and brochure, a copy of the advance letter, the press release on the PIAAC survey, as well as a folder with various newspaper articles that they could show to respondents.

3.3 Survey Operations

We will now present central aspects of the German PIAAC survey operations. First, the organization of the fieldwork, fieldwork procedures, and monitoring activities, including interview validation will be described. Second, the fieldwork outcomes and response rates will be reported.

18 Interviewers were also issued callback cards with the same "PIAAC-design" as the flyer and the brochure, to trigger familiarity and/or recognition in the target persons. This was not standard for the survey organization.

3.3.1 Field Period Organization and Sample Releases

According to the international PIAAC schedule, the data collection was to be conducted in an eight-month period, starting 1 August 2011 and ending 31 March 2012. In Germany, the data collection period coincided exactly with these dates. Countries that included PS-TRE as an assessment domain, as was the case for Germany, were required to complete at least 5 000 cases within this eight-month time frame. In Germany, we started out with 10 240 selected cases in 320 sample points, and completed a total of 5 465 cases.[19]

As described in Section 3.1 on sampling, 32 addresses were allocated to each sample point. At the outset, each sample point was worked by one interviewer. The majority of the 129 interviewers were assigned two or three sample points. In four cases, only one sample point was allocated to an interviewer. Due to the lack of availability of experienced interviewers in a few remote areas, five interviewers had to work four sample points each.

The Fieldwork Organization

The organization of the fieldwork phases was the result of careful planning that strove to specify a design that would serve several purposes:

- Allow the sample to be monitored on a regular basis during data collection, in order to identify shortfalls or bias in the sample at an early stage. To allow a timely reaction to any such developments, interviewers had to work on *all* sample points in parallel, rather than on one sample point after the other, right from the beginning of fieldwork.
- Implement a special national measure: Interviewers were to work exclusively on the PIAAC survey in the first four weeks of fieldwork, to ensure intensive contacting during the initial phase.[20]
- Optimize interviewers' workload at the beginning of the field period, in order for them to: (a) derive maximum advantage from this one-month period, in which they worked exclusively for PIAAC; (b) consolidate and internalize the project-specific skills they had trained previously; and (c) benefit from their motivational spirit right after the training.
- Split up the addresses within a sample point into two main working phases. Assure that the workload in the second main working phase is large enough for it to be financially attractive for the interviewer, and that the amount of work remains manageable, because the first re-issue of difficult cases (i.e., refusal conversion) was to start in parallel with the second main working phase.
- Intensify the level of productivity in the first main working phase in order to provide an early indication of the sample yield development and to allow time for the release of a reserve sample, if necessary.
- Undertake additional efforts and allocate sufficient time for re-working difficult and hard-to-contact cases, and locate persons not residing at the address listed in the registry. In particular, tracing these persons through a registry inquiry (*address search*; see below) is time-consuming.

A fieldwork design that took these elements into account was then established. It comprised two main working phases and several re-issue phases (see Figure 3.1). In the first main working phase, which lasted until the end of October 2011, 24 addresses per sample point were released (7 680 addresses in total). As mentioned in Section 3.2, contact letters and accompanying study materials were mailed to respondents not at one but at several dates during the first six weeks of main working phase 1. Interviewers could request one of four

19 For the definition of a completed case, see Section 3.3.4.

20 Due to the fact that all interviewers were freelancers and some worked for more than one survey organization in parallel, the exclusivity of assignments for PIAAC could not apply to assignments from survey organizations other than TNS Infratest.

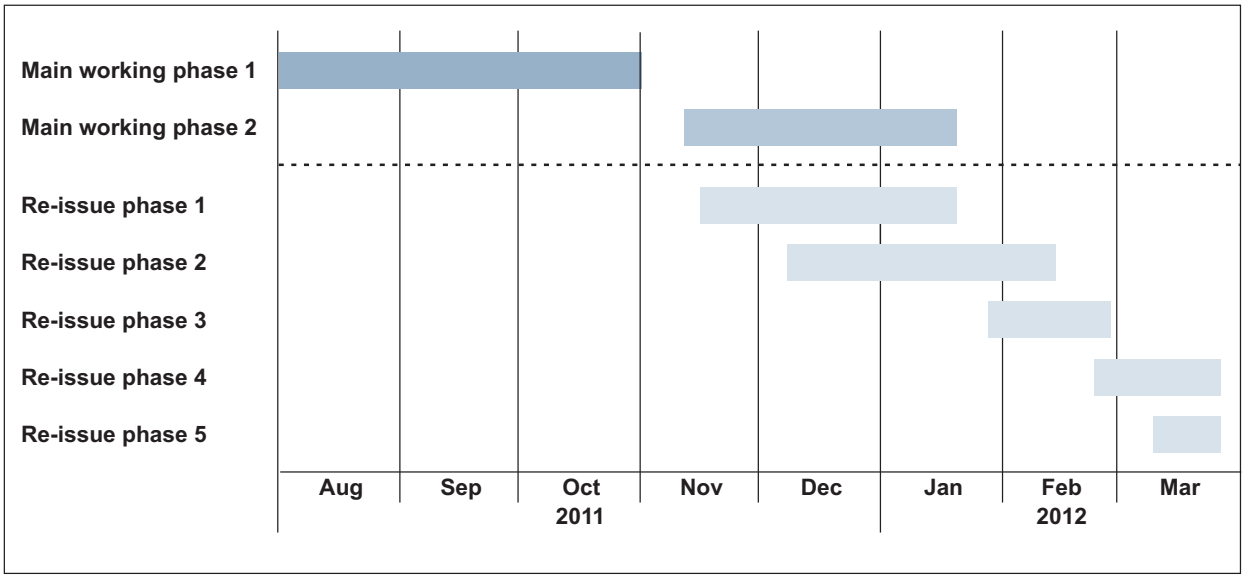


Figure 3.1. Organization of fieldwork phases

mailing dates on a case-wise basis, in accordance with their personal work plan for contacting respondents in their sample points.

Main working phase 2, with eight addresses per sample point (2 560 addresses in total), started on 11 November 2011 and ended on 20 January 2012. The first re-issue phase took place at about the same time. The eight-month data collection period allowed the implementation of five re-issue phases overall and sufficient time for several iterations of address search.

The PIAAC survey in Germany was supplemented by an additional national oversample of 26 to 55 year olds in former East Germany. This sample was collected as a part of the additional national study "Employment Opportunities of Less-educated Persons in Historical and Comparative Perspective".[21] During two phases (mid-October until end of December 2011, end of January until end of February 2012), 560 cases were additionally collected for this oversample (for more information see Annex A.3.4). Although the fieldwork for this national survey was conducted in parallel with the PIAAC main data collection in Germany, this was a separate survey and therefore the oversample data were not part of the final German PIAAC database.

Interviewer Responsibilities and Case Documentation

During the data collection period, interviewers were instructed to make at least four in-person contact attempts, as a strict minimum, before assigning a (final) disposition code and closing the case. They were to contact target persons in person on different days of the week and at different times of the day, in order to increase the chance of a contact. Furthermore, they were encouraged to invest as much additional contacting effort as possible and to be flexible when offering appointments for the actual interview.

As mentioned in Section 3.2, interviewers were equipped with a variety of documents to support them in gaining survey cooperation. Additional documents and material were necessary for conducting the PIAAC interview, for example, a show card booklet, paper exercise booklets, or material to be used by the respondent during the assessment (ruler, calculator, photos, note pads and pens). In order to be well prepared for the field, interviewers were provided with a checklist of these documents and materials (see Annex A.3.5).

21 Information on this supplementary survey can be found under http://www.wzb.eu/en/research/education-work-and-life-chances/skill-formation-and-labor-markets/projects/erwerbschancen-gering-qual (retrieved 14 March 2014).

As a rule, the PIAAC interviews were administered at the respondent's home. However, to cater for respondents reluctant to let an unknown person into their homes, interviewers were also advised to organize alternative locations for the interview, with the constraint that these were to be quiet and private enough to conduct the interview. As a supportive measure, adult education centers were centrally contacted and asked whether they could make rooms available for interviews, if the need should arise. Interviewers were provided with the corresponding contact details.

Interviewers were responsible for several tasks related to the administration of the PIAAC interview itself. They first had to administer the CAPI-based background questionnaire; this was a standard task for an experienced interviewer. Administering the assessment was, however, new to the interviewers. Although respondents worked on the assessment tasks on their own and at their own pace, the interviewer had to carry out a number of administrative tasks.

For quality control purposes, interviewers were instructed to audio-tape two full interviews early in the data collection period, provided that respondents gave their permission. Given that interviewer involvement during the assessment was minimal, these tape recordings mainly provided feedback on the interviewer's administration of the background questionnaire.

In addition, interviewers were required to fill out a contact protocol, the so-called PIAAC case folder. The case folder was a 4-page document printed on both sides of a DIN-A3-format sheet but folded to DIN-A4-format (Annex A.3.6 shows the case folder for the main working phases, in German).

The questions on the first page of the case folder were of special importance, because these data were available for both respondents and non-respondents. Information about the target person's dwelling was collected: (a) presence of an intercom, (b) type of dwelling, and (c) dwelling condition. Furthermore, two questions required the interviewers to subjectively assess the target person's education and social class. Interviewers were under strict instructions to fill out all questions on the first page (especially the two just mentioned) the first time they were at the address in question, before they attempted the very first contact, so as to have equivalent data for interviews as well as non-interviews (including non-contacts). Obviously, data obtained from the last two questions had to be treated with caution, as interviewers were asked for their "best guess" and they had little information on which to base their judgment.

The second page of the case folder was used to document contacts and contact attempts. Up to 12 contacts or contact attempts could be recorded here, with the following information:

- day of the week,
- date,
- time,
- type of contact (in-person, personal but via intercom, phone, etc.),
- disposition code,
- notes/comments,
- final disposition code, and
- for literacy-related non-respondents: age and gender as recorded at-the-door by interviewer (for more information see Section 3.3.4).[22]

22 Following the PIAAC standards and guidelines, literacy-related nonrespondents (e.g., persons who did not participate for language-related reasons) were counted as completed cases if their age and gender could be validated by the interviewer. For pragmatic reasons, age and gender were also validated for some nonliteracy-related disposition codes. This information was later discarded.

Page 3 of the case folder contained the full list of disposition codes used in PIAAC Germany (see Table 3.2). When interviewers opened the case folder, they could look at pages two and three at the same time to consult the extensive disposition code list when entering their contacting results. The list of disposition codes was based on the international disposition codes for PIAAC, with some additional national codes. At the end of fieldwork, all national codes were recoded into international codes.

Table 3.2. National PIAAC Case Folder Disposition Codes

Disposition codes
1 Completed interview
3 Breakoff
4 Refusal sample person
5 Refusal other person
6 Refusal due to time constraints
7 Language problem
8 Reading and writing difficulty
9 Learning/mental disability
12 Hearing impairment
13 Blindness/visual impairment
14 Speech impairment
15 Physical disability
16 Other disability
17 Other reasons (unspecified), such as sickness or unusual circumstances
18 Death
21 Non-contact with household
22 Non-contact with sample person
24 Sample person temporarily absent/unavailable during field period
31 Sample person moved into institution
32 Sample person moved outside country
33 Sample person moved to other community (within Germany)
34 Sample person moved, new residence unknown
35 Invalid address
Interim codes
41 Appointment
42 Sample person contact, but no final result
43 Sample person moved within community (search address and attempt contact)
44 Paused interview
90 Technical problem

The fourth page of the case folder included some questions on completed cases and breakoffs, use of paper booklets in the assessment, whether the interview was audio-taped and whether someone provided translation support during the administration of the background questionnaire.

Interviewers were required to enter some of the information from the case folder (e.g., the final disposition code) into the case management system, preferably in real time, in order for it to be available for fieldwork monitoring purposes (see also Section 3.3.2).

The case folder was modified slightly for the re-issue phases. For example, a question was included that attempted to validate work in the main working phase during refusal conversion. Furthermore, notes from the previous attempts relevant to refusal conversion were printed on the case folder.

At the end of each main working or re-issue phase, interviewers had to send their case folders, as well as the completed paper booklets, back to the survey organization's home office. In addition, at the end of data collection, all spare materials (e.g., unused paper booklets, brochures, advance letters) had to be returned.

Criteria for Re-Issues

Whether or not a case was eligible for a re-issue depended on the final disposition code the interviewer reported in the case management system during the main working phases. Not all non-interview codes could be followed up. In general, a case was considered for re-working when one of the disposition codes listed in Table 3.3 was assigned. Based on the final disposition code reported at the end of the main working phases, approximately 29% of the sample could be followed up.[23]

Table 3.3. Disposition Codes for a Re-issue

Disposition codes	
4	Refusal sample person
5	Refusal other person
6	Refusal due to time constraints
17	Other reasons
21	Non-contact with household
22	Non-contact with sample person
24	Sample person temporarily absent/unavailable during field period
33	Sample person moved to other community (within Germany)
34	Sample person moved, new residence unknown
35	Invalid address

In order to assess whether a follow-up contact was, in fact, possible, interviewer case notes were screened. In Germany, some refusals could not be re-contacted for legal reasons. Therefore, in addition, refusals based on code 4 or 5 (as indicated in Table 3.3) were scrutinized carefully to differentiate soft refusals (who may be legally re-approached) from hard refusals (who may not be re-approached). Interviewers were required to record specific reasons for all refusals according to the list given in Table 3.4 (they could mark all applicable categories); interviewers categorized refusals by sampled persons as well as proxy refusals (another household member refused for the sampled person). Cases coded with 4 (privacy issues), 6 (data confidentiality), and 8 (no participation in surveys on principle), or for which the interviewers noted these reasons in the free text entry, were *not* released for a follow-up contact. At the end of the main working phases, 27% of the sample had refused with code 4 or 5 and were subject to this screening; of these, approximately 37% (or approximately 10% of the entire sample) could subsequently be released for follow-up activities. When cases that refused with code 6 and were eligible for a follow-up were added, the rate increased to 13% of the entire sample.

23 Of these, approximately 13% were refusals (codes 4–6), 9% some type of non-contact (codes 21 and 22), 6% address-related reasons (codes 33–35), and 1% other reasons.

Table 3.4. Reasons for Refusal

Refusal codes	
1	Not interested
2	No knowledge about topic/topic too difficult
3	No time
4	Interview intrudes into privacy/won't reveal personal information
5	Participated too often in surveys
6	Data confidentiality concerns
7	Interview too long
8	No participation in surveys on principle
9	Free text entry

Address Search

When a registry-based sample is drawn, researchers are confronted with the fact that not all selected addresses are actually up-to-date when an interviewer contact is attempted. This is due to several reasons:

- variations in the degree of quality and maintenance of registry data between municipalities, and/or
- sampled persons have moved since sample selection and information has not been appropriately registered/deregistered (sampled person did not inform registry or lack of interaction between registry offices).

In other large-scale German social surveys with registry-based samples, such as the German General Social Survey (see Wasmer et al., 2012), cases with address-related disposition codes are considered to be ineligible. In PIAAC, the situation was not as simple and these cases could *not* be treated as ineligibles. However, the international Consortium allowed countries with a registry-based sample to treat persons who had moved to unknown locations or inaccessible places within the country (including municipalities that were not selected as primary sampling units during sample selection and where no PIAAC interviewers were available) as exclusions, as long as the maximum noncoverage rate of 5% was not surpassed (Mohadjer, Krenzke, & Van de Kerckhove, 2013a).

Countries were required to try to find the new location of a sampled person if the sampled residential address was incorrect or if the person had moved. If the person had moved to another PIAAC primary sampling unit or to an address where it was possible to send an interviewer, the case was to be released again. In Germany, for cases that were reported as address-related non-contacts by the interviewers,[24] meaning that the target person had moved to an unknown address and could therefore not be reached, the relevant registry office was contacted and an address search started. The address provided by registries was no longer valid for approximately 8.5% of the sample worked on during the main working phases. Approximately 1% was found to be out of scope (e.g., moved outside country). Another 1.4% did not need to be re-sent to the registries for address search because the new address had already been provided by the post office when the advance letter was undeliverable and thus returned to TNS Infratest. In 0.4 % of the cases, the information provided by interviewers was not explicit enough to conclude that the sampled address was incorrect and thus no address search was started. A total of 5.7% of the sample was sent to address search after the main working phases (583 cases). Results of this address search are given in Table 3.5.

24 The decision was based on the disposition codes after the main working phases.

Table 3.5. Results of Address Search

Result of address search	% of all cases with address search
New address	59.3
Old address confirmed	28.5
Old address confirmed, with supplement	7.2
Unknown whereabouts	1.5
No search result/sample person not registered	1.2
Moved outside country	1.7
No disclosure	0.5

Note. Percentage is based on 583 cases.

A new address was obtained for the majority of cases that were released to the address search (approx. 60%). In 36% of the cases, the old address was confirmed, and supplementary information was provided for one-fifth of these. All these cases were subsequently re-released. Table 3.6 shows the final disposition codes, after re-working, for cases for which the registry provided new addresses and for cases in which the old address was confirmed.

Table 3.6. Final Disposition Codes of Cases With Address Search Results

Final disposition code	New address (%)	Old address (%)
	n = 346	*n* = 208
Interview	39.9	13.0
Refusal	28.6	15.4
Non-contact	8.9	4.8
Ineligible	0.6	2.4
Health-related problem	1.2	1.4
Address-related problem	14.4	58.2
Other	6.4	4.8

Note. *n* = number of cases.

Address search proved to be useful in improving the response rate: Amongst cases with a new address, 40% participated in the survey. Of those for which the old address was confirmed, 13% were re-contacted and successfully interviewed.

The international Consortium required countries to flag cases that were ineligible or cases that could not be contacted due to address-related reasons and to classify them as *out of scope* or *inaccessible* (see Table 3.7).

Table 3.7. Status of Ineligibles and Status of Non-Contacts Due to Address-Related Reasons

Categories	Status	%
Deceased	Out of scope	0.2
Moved outside country	Out of scope	0.8
Moved inside country		
- Moved into institution	Out of scope	0.5
- To PIAAC PSU	Inaccessible (unknown or invalid address)	0.3
- To non-PIAAC PSU	Inaccessible (inability to interview outside PIAAC PSUs)	0.2
- To unknown PSU	Inaccessible	1.2
Unknown whereabouts	Distributed between "out of scope" and "inaccessible" categories	0.8
Invalid address	Inaccessible	0.2

Notes. Percentages of gross sample are indicated. Categories and status as defined in Mohadjer et al. (2013a, p. 9). PSU = primary sampling unit.

After several iterations of address search and follow-up contacts, overall only approximately 2% of the sampled cases were, in Germany, classified as inaccessible, according to this scheme.[25]

3.3.2 Fieldwork Monitoring and Follow-Up Strategies

We will now present information on fieldwork monitoring activities and follow-up efforts aimed at assuring the quality of survey operations, achieving high response rates, and reducing nonresponse bias. In Germany, these activities were motivated by quality control processes and procedures defined both at the international and national level.

International Quality Assurance and Control Forms

To monitor the implementation of the survey operation standards in each country, a comprehensive set of quality control checks was put in place by the international Consortium. To keep the international Consortium informed about the progress of data collection and sample development, countries systematically captured their information in custom-made electronic forms. Progress was bilaterally discussed in periodic phone conferences. Countries were provided with assistance or suggestions for improvements, if necessary. Submission dates for all forms and phone conference appointments were scheduled in advance, to meet the PIAAC timeline (Mohadjer et al., 2013a, 2013b; Montalvan & Lemay, 2013a, 2013b). Key forms during the data collection period were the *Sample Monitoring Form* and the *Quality Control Form*.

The Sample Monitoring Form was an Excel template for collecting information, on a regular basis, about the number of completed cases, response rates, and the expected yield.[26] Additionally, halfway through data collection, classification tree analyses had to be performed, in order to identify subgroups with low response rates.

The Quality Control Form was a PDF file that covered seven main sections: timeline, staff hiring and management, fieldwork, survey field management system, production and response rates, quality control of fieldwork, and ethics/confidentiality. The forms had to be submitted on a monthly basis and contents of special interest or with a significant impact were discussed in bi-monthly conference calls.

25 These cases were counted as exclusions, and thus contributed to the noncoverage rate, as described in Section 3.1.

26 Seven such forms were submitted.

National Monitoring Activities

A number of monitoring activities was carried out at the national level, both by TNS Infratest and the German National Center. Data for monitoring activities were derived from a variety of sources, such as the case management system, the interview, audio recordings (from selected interviews), and the validation process.

Interviewers used the case management system to start and close interviews, record final dispositions for non-interviews, and enter additional case information from the case folders. They were instructed to provide information on final cases as soon as possible, preferably in real time, and on a regular basis. Their data were combined with sample and further case-related information (e.g., registry data) at the home office. Throughout the data collection period, file dumps were extracted weekly or bi-weekly from this database and provided to the German National Center. These data, as well as the interview data, were then processed and reviewed for the following purposes:

- calculating response rates,
- obtaining information on non-interviews and deciding whether they could be re-issued,
- assessing interviewers' performance and flagging necessary interventions, where applicable,
- reviewing survey data (e.g., data frequencies, open-ended responses, missing data patterns),
- validating (e.g., interview duration, consistency checks with registry data, tape recordings), and
- analyzing nonresponse bias and benchmarking frequencies of central survey variables to Microcensus data.

Throughout the data collection period, TNS Infratest monitored and supervised interviewers closely and continuously discussed fieldwork progress with the German National Center. Interviewers whose performance was unsatisfactory were contacted by their supervisors,[27] in order to assess and resolve potential problems. The three least productive interviewers were taken off the PIAAC project at some time after the first main working phase. Approximately 14 interviewers received individual re-trainings or refreshers after certain issues regarding their interviewing techniques were detected (e.g., from audio recordings). Interviewers were regularly updated throughout the field period about additional fieldwork-related information (e.g., release of main working phase 2), via written memoranda and instructions.

Follow-up Efforts

Monitoring activities and results provided reliable data about response rates and nonresponse bias. Follow-up efforts were made to re-work soft refusals and non-contacts. The strategies and adjustments to fieldwork procedures listed below were implemented after the main working phase, in order to increase respondent participation.

- *Reassignment to other interviewers:* In about 18% of the cases that were re-issued during one of the re-working phases, the respondent was re-contacted by a different interviewer. Some reassignments were inevitable because respondents had moved to another location or interviewer workloads had to be re-distributed. In other cases, however, the interviewer reassignment was a strategic measure to counteract weak interviewer performance and to stimulate survey cooperation. Interviewer reassignment was most common for the following three main phase dispositions: non-contact (approx. 38%), refusal (approx. 35%), and address-related reason (approx. 26%).[28]

27 There were eight supervisors and two field managers for PIAAC.

28 In comparison: Of the re-issued cases without interviewer reassignment, approximately 46% were refusals, 31% non-contacts, and only approximately 18% were address-related non-interviews.

- *Full-timers:* Response rate development in Berlin was slow. In order to support the Berlin interviewers, a very productive interviewer from another region was sent to Berlin for a few weeks to work full-time for PIAAC. Another interviewer working in a neighboring region also supported the local interviewers during certain time periods. Both interviewers completed a substantial number of interviews in Berlin.
- *Special funding:* For re-issued cases, interviewers had access to funding for discretionary (symbolic) incentives as "door-openers". For refusal conversion and for especially difficult areas, some interviewers were offered a per diem rate as compensation.
- *Tailored advance letters:* For the re-issue phases, four tailored letters were drafted for: (a) non-contacts, (b) foreign residents, (c) low-educated respondents,[29] and (d) all others. Target persons who had moved received the standard advance letter used in the main phase.
- *Information material for immigrants:* Immigrants or respondents with restricted knowledge of the German language are often more reluctant about participating in surveys. To support interviewers in gaining cooperation of respondents with an immigration background during the re-issue phases, various additional documents were at their disposal. These included an endorsement letter prepared by the Federal Ministry of Education and Research, emphasizing that PIAAC is a serious survey of official interest, the advance letter, and an additional FAQ document; the last two documents were translated into English, Polish, Russian, Serbo-Croatian, and Turkish.

3.3.3 Validation Procedures for Fieldwork Quality Control

Validation back checks are generally considered to be one of the most important quality control features and the international Consortium defined very strict validation standards. These international standards and the validation process in Germany will now be described.

International Validation Standards

As part of the PIAAC quality control process, interviewers' work was validated to (a) verify whether an interviewer interviewed a sampled person according to survey standards, and (b) detect potential falsifications. The Consortium's criteria for validation were (OECD, 2010b):

- validation of at least 10% of each interviewer's workload,
- validation of all dispositions (completes, ineligibles, non-contacts, and refusals),
- random selection of validation cases, and
- validation conducted by supervisors, either over the phone or in person.

According to the standards, if falsification was detected, or if a case was found to be suspect, all finalized cases of that particular interviewer had to be validated. Falsified cases had to be re-worked by another interviewer. As mentioned earlier, interviewers had to audio-tape at least two full interviews. These audio recordings also had to be reviewed.

Interview Validation Process in Germany

The survey organization's standard validation practice for high-quality surveys differs from that specified in the international validation requirements. It encompasses: (a) a 100% validation of completes, i.e., a validation questionnaire is sent to each respondent, (b) consistency checks between interview and registry data (age, gender, citizenship), and (c) phone validation for cases with inconclusive or inconsistent questionnaire information. The focus

29 Because information about education was not provided by registries, educational status was imputed for this purpose, using paradata.

of the German standard validation strategy is on the identification of falsified interviews. In general, there is little incentive for interviewers to falsely report cases as refusals or non-contacts, because they only receive payment for completed cases.

The German PIAAC validation strategy was based on the national standard approach, but with various additional measures and efforts to extend the validation process according to the international standards.[30] Complying with the requirement of validating all dispositions posed some challenges because, for example, German data protection laws prohibit a re-approach of hard refusals. Non-standard additional validation measures implemented in PIAAC Germany were:

- *Validation of some ineligibles*: This included, for example, Internet search of institutions (for target persons who had been institutionalized).
- *Validation of refusals due to disability by mail:* A separate validation questionnaire was developed for these dispositions.
- *Validation of non-contacts by phone:* This procedure was carried out in the last month of data collection, when additional contact attempts by interviewers were not successful.
- *Validation of soft refusals in person:* When a soft refusal was re-worked by an interviewer other than the one in the main working phase, this second interviewer was instructed to collect additional information about whether the respondent had been previously contacted by an interviewer. The interviewer had to record this information in the case folder for the re-working phases.

In addition to reviewing audio tapes for quality assurance purposes, interview length and consistency of specific interview contents were checked on a regular basis.

Results of National Interview Validation

About 56% of the German gross sample of 10 240 cases was selected for validation. The majority of these cases, 92%, were completes. These underwent a registry data check and respondents were sent a validation questionnaire; 63% returned a completed validation questionnaire to the survey organization. Those that did not return the questionnaire were not classified as "successfully validated", according to international validation standards.

Of the above-mentioned cases that were selected for validation, approximately 62% were successfully validated either by mail, Internet search, phone, or in-person.[31] However, among the remaining 38% that were not successfully validated by one of these methods, 87% were completes that were at least successfully validated by the registry data check. With respect to the international requirements, German validation results were as follows:

- *Validation of at least 10% of each interviewer's workload:* At least 10% of each interviewer's workload was successfully validated (minimum: 11%, maximum: 52%, average: 35%). Because validation focused on completed interviews, the number of validated cases varied due to different performance rates.
- *Validation of all dispositions:* There are several reasons (stated above) why the validation of all dispositions was not fully feasible in Germany. However, an attempt was made to extend the validation process by including other dispositions, albeit on a smaller scale. Overall, approximately 5% of the cases that were successfully validated had non-interview dispositions.
- *Random selection of validation cases:* Random selection of cases was not generally suitable for the German validation process. First and foremost, validation in Germany focuses on the validation of *all* completes. In addition, *all* sampled persons who refused to par-

30 For more detailed information on control and validation of interviewers' work in PIAAC Germany, see Massing, Ackermann, Martin, Zabal, and Rammstedt (2013).

31 Thus, 35% of the gross sample was successfully validated.

ticipate due to disabilities were selected for validation. Thus, the validation of all completes and all non-interviews due to disability exceeded the international random selection requirement. The only refusals that were eligible for validation in Germany were soft refusals that could be legally re-approached; they were validated if a different interviewer was assigned to re-work the case. Only respondents who had moved into an institution were selected at random.

- *Validation to be conducted over the phone or in person:* As mentioned earlier, standard procedures are registry data check and mail validation questionnaires. In order to expand the spectrum of validation modes, some of the successfully validated cases were validated by Internet search (1%), by phone (5%),[32] or in person (2.5%).

A total of 258 audio recordings was expected from 129 interviewers. Eight interviewers did not submit any tape recordings. Of the remaining 121 interviewers, the majority (116) submitted two recordings each, four submitted three recordings each and one interviewer returned only one recording. Of these 245 audio recordings, the German National Center reviewed the *full interview* (background questionnaire and assessment) for 31%, and the *background questionnaire only* for 69%. As a result of the audio-tape reviews, a few interviewers were provided with re-training to address specific issues with their interviewing technique.

Validation of interviews revealed isolated cases in which an interviewer had unintentionally conducted the interview with a wrong person. These cases were re-issued. The validation and monitoring activities provided no evidence of any falsification.

In sum, the quality control measures implemented to monitor interviewers' work seemed to have been effective. Considerable effort was put into a thorough validation of interviews through registry data check, respondent questionnaires, and follow-up phone calls. The audio-tapes were very informative regarding the adherence to standardized procedures and interview quality. However, reviewing the tapes was also very time-consuming. The amount of information gained from the additional, innovative features of validation was limited, especially considering the amount of effort they required.

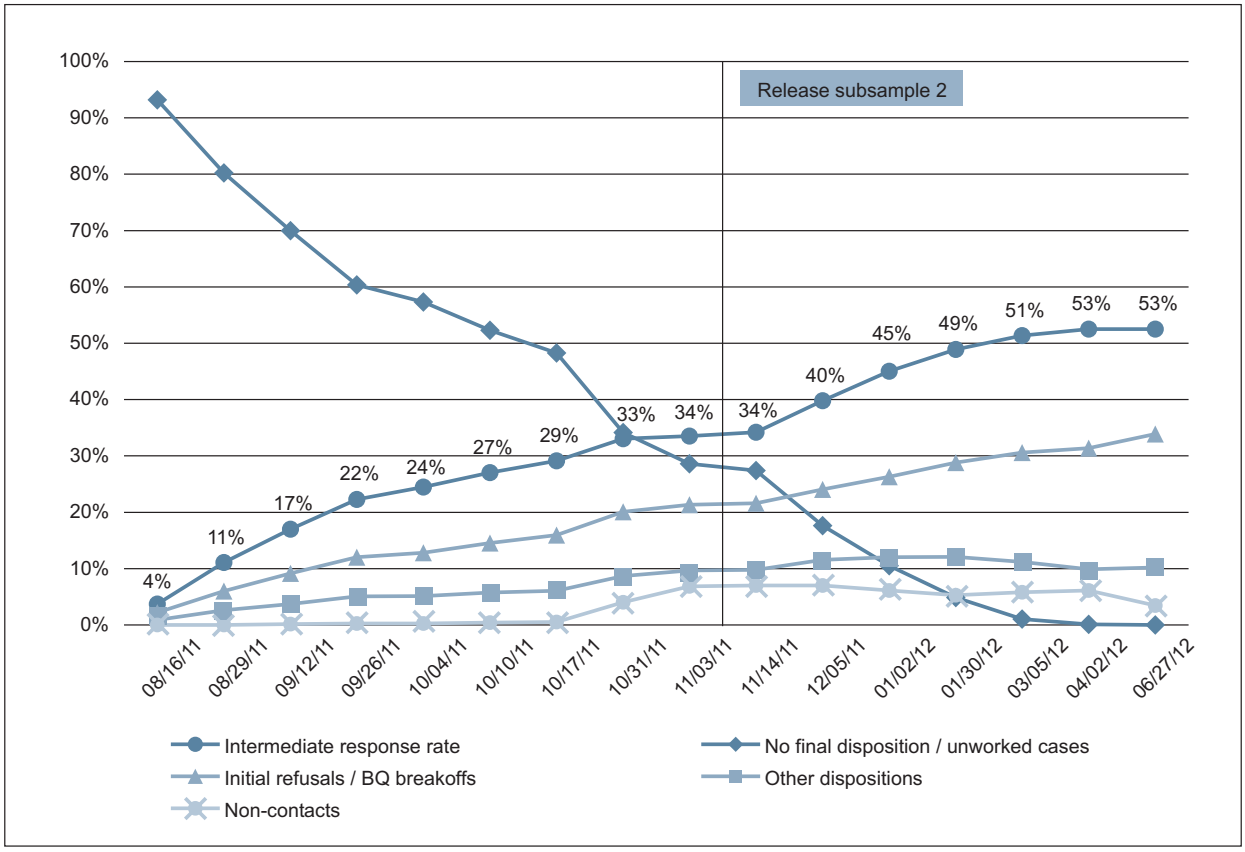
3.3.4 Fieldwork Results

During the data collection period, the development of the response rate, as well as of other dispositions, was closely monitored by TNS Infratest and the German National Center. To this end, a temporary fieldwork indicator for the response rate was computed:

$$\textit{Fieldwork response rate} = \frac{\textit{Number of cases with completed background questionnaire}}{\textit{Gross sample of } 10\,240 \textit{ cases}} \quad (1)$$

The other dispositions that were monitored over time were (see Figure 3.2): (a) initial refusals and background questionnaire breakoffs, (b) non-contacts (see definition below), (c) other dispositions (e.g., disabilities, ineligibles), and (d) unworked cases or cases without a final result. Figure 3.2 also displays results at the end of June 2012, although data collection was finalized by the end of March 2012. The information provided for June 2012 is based on cleaned data and represents the final German fieldwork results based on the fieldwork response rate indicator.

32 In general, much more phone validation than usual was performed for PIAAC.



Note. Reported results are unweighted and cumulated. Response rates refer to temporary fieldwork results. Other dispositions were classified into four key groups. The number of cases per group was divided by 10 240 (gross sample) for each reporting date. BQ = background questionnaire.

Figure 3.2. Response rate development and dispositions in PIAAC Germany over fieldwork period

The response rate increased considerably during the first six to eight weeks of data collection. At this time, interviewers worked exclusively for PIAAC, and it is likely that the more cooperative target persons were interviewed. By October 2011, the response rate reached approximately 30% and then stagnated until the release of new cases in mid-November 2011. Field efforts in the main working phase 2, as well as during the re-issue phases, accounted for a continuous rise of the response rate throughout the remaining fieldwork period. The rate of unworked cases is almost inversely related to the response rate development. The non-contact disposition was not reported back by interviewers until the end of October 2011, when interviewers were instructed to finalize the cases of the first main working phase. The non-contact rate dropped slightly during the first quarter of 2012 but, overall, remained fairly constant during fieldwork. It then declined during the final cleaning process, as described below. Figure 3.2 clearly shows that the most common reason for a non-interview is a refusal.[33] Despite intensive refusal conversion efforts, the refusal rate increased continuously over time.

Before we report the final fieldwork results according to the international response rate definition in more detail, it is necessary to specify what is regarded as a completed case in PIAAC.

33 Breakoffs in the background questionnaire were negligible.

Definition of a Completed Case

In PIAAC, a completed case is not simply a case with a finalized interview (i.e., respondent answered the background questionnaire and completed the cognitive assessment). In fact, in certain instances, incompletes due to refusals or breakoffs at different stages and for a variety of reasons are regarded as completed cases in PIAAC. The international Consortium defined a completed case in the standards and guidelines (OECD, 2010b). According to this definition, a case is completed if one of the two following conditions applies:

- The case contains at least responses to central background questions[34] and the core (see Section 2.1) was administered (including cases with incomplete or missing core data, if the core was broken off due to literacy-related reasons).
- The sampled person is a literacy-related nonrespondent (LRNR), for whom age and gender were collected by the interviewer.

Literacy-related reasons for non-interviews or breakoffs are (a) language problems, (b) reading and writing difficulties, and (c) learning or mental disabilities. The LRNRs, in particular, are sampled persons who were not capable of completing the PIAAC background questionnaire for one of these reasons. If interpreters were not available to support the administration of the background questionnaire, interviewers were instructed to obtain at least information on age and gender (OECD, 2013a). The LRNRs for whom this basic information was obtained were treated as part of the net sample. The international Consortium considered it important to represent the LRNRs in the net sample, because they are likely to differ from other survey respondents with regard to their competencies (Mohadjer et al., 2013c). In its Skills Outlook Report, the OECD (2013a) reports that these persons most likely only have lower levels of proficiency and that, in the majority of countries, they represent less than 5% of the population.

Of the 10 240 sampled persons in Germany, 131 were initial LRNRs, and of these, age and gender information was collected for 86. These 86 LRNRs, plus 5 379 survey respondents with background information—according to the definition given above—formed the realized net sample of 5 465 completed cases.

Final Response Rates and Disposition Codes for PIAAC Germany

The calculation of the overall response rate is based on the distinction between background questionnaire and assessment, on the one hand, and different reasons for non-interviews, on the other. The *overall response rate* for countries with registry samples is the product of the *background questionnaire response rate* and the *assessment response rate* (Mohadjer et al., 2013a). Table 3.8 specifies the equations used to compute both the background questionnaire response rate and the assessment response rate for registry countries, such as Germany, as defined by the international Consortium.

34 Minimum requirement: age, gender, highest level of education, and employment status.

Table 3.8. International PIAAC Response Rate Calculation for Registry Countries

Stage	Response rate calculation	Description
Background Questionnaire (for countries with registries)	COMPLETE / (ELIGIBLE - EXCLUDE)	C^b = Completed BQ cases
	COMPLETE = $C^b + LR^b$	LR^b = Literacy-related nonrespondents
	ELIGIBLE = $SP^b - D^b - I^b - U^b * ((D^b + I^b) / K^b)$	SP^b = All sampled persons
	EXCLUDE = ELIGIBLE * EXC_PROP	D^b = SPs with a disability
		I^b = SPs known to be ineligible
		U^b = SPs with unknown eligibility status
		K^b = SPs with known eligibility status
		EXC_PROP = Inaccessible rate
Assessment[1]	COMPLETE / ELIGIBLE	C^a = Completed assessments
	COMPLETE = $C^a + LR^a$	LR^a = Literacy-related nonrespondents
	ELIGIBLE = $C^b - D^a - I^a$	C^b = Completed BQ cases
		D^a = SPs with a disability
		I^a = SPs known to be ineligible

[1] The assessment response rates with and without reading components were computed using the same formula, the difference being reflected in how each SP was classified, whether completing the reading components or not.

Notes. Source: Mohadjer et al. (2013a, page 11, extract of Table 16-3). BQ = background questionnaire.

Thus, for the German *background questionnaire response rate* calculation, the numerator consists of the cases that are completed cases, according to the definition given above (5 465). The denominator consists of all sampled persons (10 240) minus the following groups of persons:

- Persons with disabilities, for example, hearing or visual impairment (disposition codes 12, 13, 15, or 16). In principle, these persons are eligible, but the assessment is not designed in a way that accommodates their particular circumstances (Mohadjer et al., 2013a).
- Ineligibles, for example, deceased persons (disposition codes 18, 31, or 32).[35]
- A certain proportion of persons with unknown eligibility status, such as persons who moved to another location but could not be traced.
- A very small number of eligible persons that was excluded from the response rate calculation (such as inaccessible persons, see above).

The *assessment response rate* is calculated as the number of completed assessment cases divided by the number of eligibles for the assessment. Completed cases, in terms of the assessment, are respondents who either finalized the assessment or who broke off for literacy-related reasons; in Germany, the numerator for the assessment response rate calculation was 5 337 cases, of which 23 were LRNRs and six were cases with technical problems. Eligible respondents are cases with a completed (or almost completed) background questionnaire minus persons with a disability and ineligibles.[36] In a few instances, interviewers unintentionally misplaced a paper booklet and thus the assessment data were missing for the respective respondent. Cases with missing booklets were treated as nonrespondents for the assessment response rate calculation.

35 Codes 31 and 32 are national codes that were recoded into the international code 25.

36 In Germany, there are 5 374 eligible respondents (5 379 cases with a completed background questionnaire minus five cases who broke off the assessment, due to a disability).

The international Consortium reported response rates that are comparable across all countries, based on the final count of completed assessments without reading components, because this option was not implemented by all countries. For the calculation of the assessment response rate, it is therefore irrelevant whether a respondent successfully worked through all the reading components tasks or not. To account for unequal selection probabilities, the response rates reported by the international Consortium were weighted by the country's design weight. The overall design-weighted German PIAAC response rate, as computed by the international Consortium, was 55% (Mohadjer et al., 2013a, p. 12). Table 3.9 shows the unweighted and design-weighted response rates for Germany:

Table 3.9. German PIAAC Response Rates

Response rate	Unweighted (%)	Design weighted (%)
BQ	55.5	55.3
Assessment	99.3	99.3
Overall	55.1	54.9

Notes. The assessment response rate is computed without reading components and cases coded as "missing booklets". BQ = background questionnaire.

Table 3.10 shows the national final distribution of disposition codes for the unweighted German gross sample. These final disposition codes are based on (a) the partially more differentiated national codes from the case folder, (b) the final status of cases that were administered the reading components, and (c) results *including* cases with missing booklets as completed assessments.[37]

37 Contrary to the international definition, at the national level, the cases with missing booklets were counted as completed assessments, because the entire interview was, in fact, carried out, the respondent received the incentive, and the interviewer was paid. The fact that the booklet accidentally went missing was not relevant.

Table 3.10. Final National Disposition Codes for PIAAC Germany

Final disposition code	*n*	%
1 Completed interview	5 319	51.9
3 Breakoff[1]	8	0.1
4 Refusal sample person	2 980	29.1
- Initial	2 974	
- In assessment	6	
5 Refusal other person	237	2.3
- Initial	236	
- In background questionnaire	1	
6 Refusal due to time constraints[2]	255	2.5
7 Language problem	114	1.1
- Initial, with age/gender collected	60	
- Initial, without age/gender collected	38	
- In assessment	16	
8 Reading and writing difficulty	17	0.2
- Initial, with age/gender collected	3	
- Initial, without age/gender collected	1	
- In assessment	13	
9 Learning/mental disability	30	0.3
- Initial, with age/gender collected	23	
- Initial, without age/gender collected	6	
- In assessment	1	
12 Hearing impairment[2]	5	0.0
13 Blindness/visual impairment	7	0.1
- Initial	5	
- In assessment	2	
14 Speech impairment[2]	5	0.0
15 Physical disability	13	0.1
- Initial	9	
- In assessment	4	
16 Other disability[2]	17	0.2
17 Other reasons (unspecified), such as sickness or unusual circumstances	158	1.5
- Initial	154	
- In assessment	4	
18 Death[2]	24	0.2
21 Non-contact with household[2]	351	3.4
22 Non-contact with sample person[2]	211	2.1
24 Sample person temporarily absent/unavailable during field period[2]	72	0.7
31 Sample person moved into institution[2]	47	0.5
32 Sample person moved outside country[2]	83	0.8
33 Sample person moved to other community (within Germany)[2]	51	0.5
34 Sample person moved, new residence unknown[2]	175	1.7
35 Invalid address[2]	55	0.5
90 Technical problem[1]	6	0.1
Total	**10 240**	**100.0**

Notes. In two cases, a respondent broke off the background questionnaire at a late stage. These cases were treated as completed cases because the minimal background questionnaire requirements were met, and they are therefore part of the net sample (and were assigned a final disposition code of 17 in the assessment). *n* = number of cases.

[1] Only assessment. [2] Only initial.

As previously mentioned, interviewers had to make at least four attempts to contact the target person. In Germany, 4.5 contact attempts were made, on average, before a final disposition was obtained (compared to an average of seven contact attempts for non-contacts). The international standards required countries to keep the non-contact rate under 3% (OECD, 2010b). In order to achieve this goal in Germany, special attention was paid to additional fieldwork efforts. The vast majority of cases with a non-contact disposition at the end of the main working phase 1 were re-worked in one of the re-issue phases. In the final weeks of data collection, interviewers were instructed to invest particular effort into reducing the number of non-contacts. As a supplementary measure, the German National Center conducted final contact attempts by phone. In addition, during data cleaning, the German National Center inspected all case folders for non-contacts individually, to check whether the interviewer had recorded a contact with the household in this document, but mistakenly reported a non-contact as the final disposition in the case management system. Usually this situation occurred when an interviewer had recorded several contact attempts and at least one of those was a contact, but the last one on that list was a non-contact. Where applicable, the final disposition code was corrected. The final non-contact rate, defined as no contact whatsoever with the household (code 21), was 3.4% for Germany, only slightly above the international benchmark.

3.3.5 Interview Setting

At the end of the PIAAC interview, each interviewer had to document information about the interview setting and his or her interaction with the respondent. Information that was obtained included where the interview took place and whether the respondent was distracted by other activities or media, such as a TV.

In Germany, on average, there was another person present during every fifth interview (see data for 5 375 cases given in Annex A.3.7). On a very small number of occasions, this person helped or tried to help the respondent with the background questionnaire or the assessment, although the kind of support provided is not further specified. According to the interviewers, the majority of the respondents, approximately 90%, seemed to understand the interview questions. Almost 19% asked the interviewer for further clarification during the assessment. However, information from debriefing sessions with the interviewers indicates that many of these inquiries actually referred to the orientation module of the assessment. Sources of distraction or interruptions did not seem to be a general problem. Although 8% of respondents communicated by phone, text message, or email at some point during the interview, respondents almost never undertook domestic tasks, such as cooking, in parallel. The majority of the interviews (almost 87%) were conducted in either the living/dining room or in the kitchen. Only 4.4% of the respondents made use of the option to be interviewed in a location outside their household, such as a library.

3.3.6 Interviewer Debriefing

Following data collection, information and feedback about the PIAAC fieldwork was obtained from interviewers through an extensive debriefing questionnaire that was nationally adapted and extended from the debriefing form produced by the international Consortium. This questionnaire was completed by 93% of the interviewers. Additionally, three half-day debriefing workshops were held, each attended by approximately 15 interviewers, TNS Infratest project directors as well as other home-office staff, and members of the German National Center. Interviewers presented and discussed their experiences during fieldwork at these workshops.

Overall, the debriefing results confirmed that the PIAAC fieldwork in Germany went extremely well. The interviewers enjoyed the PIAAC work and felt privileged to have been a part of this unique survey. They were satisfied with their fieldwork results and coped well with the complexities of the PIAAC interview. Interviewers attributed part of their success with PIAAC to their training and the comprehensive and professional supplementary materials.

Respondents' reactions to PIAAC were very heterogeneous, ranging from being very interested in the topic and eager to participate, to not at all interested and afraid of having to do the assessment. When interviewers were asked to compare gaining respondent cooperation in PIAAC with that in other high-quality national registry surveys, their overwhelming response was that the 50 euro incentive was the most persuasive PIAAC-specific advantage and a crucial factor (albeit not the only one) in achieving such high response rates.[38] Additionally, the international dimension of PIAAC was regarded as a convincing argument for recruitment, as were the interesting survey topic and the relevance and importance of the survey for Germany. The most frequently mentioned survey-specific disadvantage was the "test situation", which induced caution or anxiety in a number of target persons. However, whereas the test situation was a deterrent for some, for others, the opportunity to test themselves was regarded as a challenge, and motivated them to participate. The unusual length of the interview was mentioned as another PIAAC-specific obstacle to gaining respondent cooperation.

The interviewers felt very well equipped with their extensive at-the-door material. The attractive accompanying material (flyer and brochure) was well received and helped to spark the interest of target persons. Some of the material was not universally useful but was helpful only for certain, specific cases. This was expected, because the diversity of the materials supplied to interviewers was intended to provide them with a choice of approaches for diverse at-the-door situations.

Overall, the administration of the PIAAC interview went smoothly. There were very few or no technical problems, and support from the technical hotline was very efficient. The majority of the interviewers felt comfortable with their role in the assessment. In the computer-based assessment, there were some problems with the highlighting functionality. With respect to the items for PS-TRE, respondents sometimes mentioned that they generally did not have problems doing those kinds of tasks with their own computer and computer programs, but that the email and the Internet interface implemented in the assessment were very unusual and restricted. Some respondents apparently had difficulties reading the text on the computer screen during the assessment, usually due to the small font size. A number of respondents criticized the length and textual overload of the items.

38 However, it should be noted that this high incentive also had drawbacks, because it caused skepticism and mistrust in a few cases.

3.4 Quality of the Sample Data

A central indicator of high-quality sample data is low bias in the study estimates. In theory, data from probability-based samples provide unbiased estimates of the target population. This sampling approach assumes that every eligible adult has a nonzero chance of selection, can be located, and is willing to participate in the study and respond to all survey items (Mohadjer et al., 2013a). In practice, however, sample data are subject to several sources of error that arise during the design and implementation of a survey. For example, *undercoverage* occurs when persons in the target population are excluded from the sample frame, for example because they live in locations that are inaccessible to interviewers. *Unit nonresponse* occurs when sampled persons cannot be successfully contacted or are not willing to participate in the survey. *Item nonresponse* occurs if participants choose not to respond to specific questions (Biemer, 2010).

These errors may lead to biased estimates if the excluded or non-participating individuals from the target population differ from respondents regarding characteristics that are closely related to the study outcome(s). For example, as equation (2) shows, the bias of an estimator can be expressed as a function of the population covariance between response propensity and the study outcome, in relation to the average response propensity in the target population (adapted from Bethlehem, 2002, p. 276):

$$Bias\left(\bar{y}_R\right) = \frac{\sigma_{yp}}{\bar{p}} \quad (2)$$

where: σ_{yp} = population covariance of the outcome variable and response propensity;
$\bar{p}$ = mean response propensity in the target population.

In order to minimize such survey errors, survey practitioners generally aim at keeping exclusion rates low and at reducing nonresponse during fieldwork. Furthermore, weighting is commonly used to reduce bias remaining in the data set after data collection (Gabler & Ganninger, 2010).

3.4.1 Weighting

In PIAAC, weighting adjustments were either conducted by the international Consortium or by participating countries themselves. Countries that calculated their own weights were required to follow the procedures outlined in the *PIAAC Weighting and Variance Estimation Plan* (OECD, 2011c), in order to guarantee the calculation of comparable estimates of proficiency and their sampling error across countries.

Germany commissioned the international Consortium to conduct the weighting adjustments for the German PIAAC data.[39] However, it remained a country responsibility to identify the variables to be used for weighting. This was done via several *nonresponse bias analyses* (NRBAs). In Germany, for different weighting steps (see below), variables related to age, gender, citizenship, municipality size, region, and education were selected. We will first describe the general weighting steps followed in PIAAC, and then specify how these weighting variables were selected.

39 The international Consortium conducted weighting for approximately half of the other participating countries in PIAAC.

The weighting approach in PIAAC included the following adjustment steps to reduce bias:

1) base weight,
2) unknown eligibility adjustment,
3) nonresponse adjustment,
4) trimming, and
5) calibration.

Each weighting step yielded a new weight for all or for a subgroup of the sample units, which then served as the basis for the next weighting step. At the end of this process, a final weight that allows for inferences to the target population was obtained. Table 3.11 provides details of the stepwise calculation of the weights. All cases with a completed background questionnaire (R), and literacy-related nonrespondents for whom age and gender were successfully collected (L1),[40] received a final weight (F6). As mentioned in Section 3.3, literacy-related nonrespondents are regarded as a part of the PIAAC target population that cannot be represented by survey respondents because they presumably differ from respondents in terms of their proficiency (Mohadjer et al., 2013c).

Table 3.11. Adjustment Factors and Weights

Weighting step	Factor	Weight
Base weight	N/A	$W_l = \frac{1}{P_{hl}}$
Unknown eligibility adjustment	$F_{1l} = \begin{cases} \frac{S_R + S_{NR} + S_{L1} + S_{L2} + S_D + S_I + S_U}{S_R + S_{NR} + S_{L1} + S_{L2} + S_D + S_I} & \text{if } l \in \text{I} \\ \frac{S_R + S_{NR} + S_{L1} + S_{L2} + S_D}{S_R + S_{NR} + S_{L1} + S_{L2} + S_D + S_I} & \text{if } l \in \text{U} \\ 1 & \text{if } l \in \text{R, NR, L1, L2, D} \end{cases}$	$W_l F_{1l}$
Nonliteracy-related nonresponse adjustment	$F_{3l} = \begin{cases} 1 & \text{if } l \in \text{L1, L2, I} \\ \frac{S_R + S_{NR} + S_D + S_U}{S_R} & \text{if } l \in \text{R} \\ 0 & \text{if } l \in \text{NR, D, U} \end{cases}$	$W_l F_{1l} F_{3l}$
Literacy-related nonresponse adjustment	$F_{4l} = \begin{cases} 1 & \text{if } l \in \text{R, I} \\ \frac{S_{L1} + S_{L2}}{S_{L1}} & \text{if } l \in \text{L1} \\ 0 & \text{if } l \in \text{L2} \end{cases}$	$W_l F_{1l} F_{3l} F_{4l}$
Trimming	$F_{5l} = \begin{cases} 1 & \text{if } W_l F_{1l} F_{3l} F_{4l} \le cutoff \\ \frac{cutoff}{W_l F_{1l} F_{3l} F_{4l}} & \text{if } W_l F_{1l} F_{3l} F_{4l} > cutoff \end{cases}$	$W_l F_{1l} F_{3l} F_{4l} F_{5l}$
Calibration	$F_{6l} = \frac{S^*}{S_R + S_{L1}}$ (for poststratification) See Deming and Stephan (1940) for raking adjustments and Särndal, Swenson, and Wretman (1992) for GREG estimation.	$W_l F_{1l} F_{3l} F_{4l} F_{5l} F_{6l}$

Note: The factors and weights shown here are for person *l*. The persons can be classified as R: BQ respondent who is not assessment literacy-related nonrespondent, L1: BQ literacy-related nonrespondent with age and gender successfully collected or assessment literacy-related nonrespondent, L2: BQ literacy-related nonrespondent with age or gender not successfully collected, NR: BQ nonliteracy-related nonrespondent, I: ineligible, D: sampled person with a disability, or U: sampled person with unknown eligibility status. *S* represents the sum of the prior-stage weights over records in the same adjustment cell as person *l*, and S* is the control total for the cell. *P* represents the selection probability. The factor F_2 is reserved for countries with screeners.

Notes. Source (minimally modified): Mohadjer et al. (2013c, p. 3, Table 15-1A). BQ = background questionnaire. *F* = adjustment factor. N/A = not applicable.

40 L1 includes both initial literacy-related nonrespondents with age and gender successfully collected by the interviewer (previously introduced as LRNR) as well as literacy-related breakoffs in the assessment.

Base Weight

The base weight, which is computed as the reciprocal of the person selection probability, was calculated to correct for the differential selection probabilities of each sampled person (see Table 3.11). Due to an erroneous selection algorithm at the second stage of sample selection in PIAAC Germany, person probabilities of selection were unequal, so that some persons had a greater chance of being included in the sample than others. The error and its impact on person selection probabilities are described in Section 3.1. This issue was corrected, through base weights, by giving persons with a larger selection probability a smaller base weight and vice versa.

Unknown Eligibility Adjustment

A few sampled persons could not be located, for example, due to the fact that they had moved and their new address could not be traced. The eligibility of these persons could therefore not be determined. It was possible, for instance, that a number of these persons had moved outside the country or to an institution. With the unknown eligibility adjustment weighting step, these persons were down-weighted according to the proportion of eligible cases among those with known eligibility. This procedure made it possible to treat the down-weighted cases with unknown eligibility status as eligible nonrespondents in the nonresponse adjustment (Mohadjer et al., 2013c). The adjustment was carried out within the weighting cells defined for the nonresponse adjustment (see below).

Nonresponse Adjustment

For this adjustment step, nonrespondents were divided into two groups: nonliteracy-related and literacy-related nonrespondents. The literacy-related nonrespondents did not participate in PIAAC due to reading or writing difficulties, language problems, or a learning/mental disability. The nonliteracy-related nonrespondents were simply not willing to participate, had no time, could not be contacted during data collection, or could not participate due to a physical disability or impairment.[41] Presumably, these two groups of nonrespondents differ with regard to their proficiencies. Nonliteracy-related nonrespondents are assumed to have proficiencies that are similar to those of respondents, whereas literacy-related nonrespondents are considered to be systematically different from respondents because reading, writing, and language skills are necessary for completing the direct skills assessment. For this reason, they were treated separately in this weighting step (Mohadjer et al., 2013c).

As shown in the equation for the nonliteracy-related adjustment step in Table 3.11, the combined unknown eligibility-adjusted weights for nonliteracy-related nonrespondents, sampled persons with a disability, and down-weighted unknown eligibility cases were distributed to respondents. The nonresponse adjustment was performed within cells that were defined on the basis of variables identified in the basic NRBAs.

In Germany, only a restricted number of variables (age, citizenship, municipality size) and categories were used for nonresponse weighting to ensure that a sufficient number of cases would be available in each weighting cell ($n \geq 24$). However, in the group of non-Germans, some weighting cells had to be collapsed across age groups and municipality size in order to achieve the minimum number of cases per cell.

In the literacy-related nonresponse adjustment step, the weights of the literacy-related nonrespondents to the background questionnaire without age and gender collected (L2; $n = 45$) were distributed to literacy-related nonrespondents to the background questionnaire with age and gender collected ($n = 86$) and literacy-related nonrespondents to the assessment (i.e., breakoffs, $n = 23$); the last two groups are both regarded as L1 cases here. Due to the small number of literacy-related nonrespondents, the weighting cells of the nonliteracy-related nonresponse adjustment could not be used for the adjustment. Thus, instead

41 More information on disposition codes in PIAAC is given in Section 3.3.

of distributing the literacy-related cases across the nonresponse weighting cells, they were grouped in one cell.

Calibration

In this step, the nonresponse-weighted data were matched to high-quality population data, such as Census information. This calibration was used to control for undercoverage bias, to reduce the mean square error of estimates, and to create consistency with statistics from other studies (Mohadjer et al., 2013c).

Countries could choose to use one of the three main calibration techniques: poststratification, raking, or generalized regression estimators (GREG). As a minimum, countries were required to benchmark the data to control totals for age and gender (Mohadjer et al., 2013c). In Germany, data were poststratified to match population totals for age, gender, region, and education (i.e., the highest general school leaving qualification obtained).[42] It was particularly important to include an adjustment for education because various NRBAs had shown that respondents' level of education was a strong predictor of participation. However, this potential for bias could not be addressed in the previous nonresponse weighting step, because data of sufficient quality for assessing educational attainment (e.g., from the registries) were not available for the nonrespondents.

For poststratification, population counts at the cell level were needed. In Germany, these were provided by the Federal Statistical Office. Some cells had to be collapsed to achieve the necessary minimum number of cases per cell across all weighting variables, with the exception of "region".

Trimming

After calibration, outlier weights resulting from weighting adjustments were trimmed and re-calibrated to avoid having extreme weights that add to variance inflation. Outlier weights in PIAAC Germany were mainly caused by variability in the selection probabilities. For more information on the method and criteria for trimming, as well as on the entire weighting process, see Mohadjer et al. (2013c).

After the final weights had been produced by the international Consortium, thorough weighting checks were conducted by the German National Center and GESIS sampling and weighting experts. For example, we checked whether the design effect[43] increased due to weighting and examined the distribution and range of weights. Moreover, the weighted distributions of the variables used in nonresponse adjustment and the calibration step were compared to the distribution of the variables in the benchmark datasets.[44] The analyses showed that the weighted distributions were virtually identical to the benchmark data[45] and that the weights did not show irregularities. Thus, the weights were approved by the German National Center.

42 "Region": Germany's 16 federal states were grouped into three categories: North, South, and East. "Education" includes the three main general school leaving qualifications in Germany "Hauptschulabschluss", "Realschulabschluss", and "Abitur", plus a separate category for students who are still in general school.

43 The design effect is the effect of the sampling design on the variance of an estimator.

44 The eligible sample was used as benchmark data for nonresponse weighting. Data from the Microcensus 2010 were used as benchmark for the calibration step.

45 Minor deviations of up to 0.2 percentage points were due to the fact that weighting cells were collapsed. In the calibration step, there are additional deviations, due to rounding, because the Federal Statistical Office provided Microcensus data rounded to 1 000 persons, for reasons of data confidentiality.

3.4.2 Nonresponse Bias Analyses

As mentioned in the introduction to this section, frame error (i.e., coverage error) and nonresponse error are regarded as important sources of survey error that may cause bias in sample data (Biemer, 2010). In order to minimize these survey errors, participating countries were required to keep exclusion rates low and to reduce nonresponse before, during, and after data collection (Mohadjer et al., 2013a).

With an undercoverage rate of 2.5%[46] of the PIAAC target population, bias due to undercoverage was negligible in PIAAC Germany. Efforts to reduce nonresponse before and during data collection are described in Sections 3.2 and 3.3.

In order to detect nonresponse bias after data collection, all participating countries had to conduct *basic nonresponse bias analyses* (basic NRBAs). These analyses explored factors that influenced the propensity of target persons to respond to the study and served to identify appropriate variables for the nonresponse adjustment. In addition, all countries with an overall response rate below 70%—and thus also Germany—were required to conduct *extended nonresponse bias analyses* (extended NRBAs). These extended analyses were carried out after the PIAAC weights had been produced. They aimed primarily at assessing the level of nonresponse bias remaining in the proficiency estimates after the selected weighting procedure had been applied (Mohadjer et al., 2013a).

Furthermore, *item nonresponse bias analyses* (item NRBAs) had to be conducted for all background questionnaire items with an item response rate below 85%. In PIAAC Germany, item nonresponse was, in general, very low and only three items were found to have a response rate below 85%. However, these three items referred to respondents' earnings in their current job and were a part of a *set* of items that were to be combined, in order to achieve comparable income information for as many respondents as possible. Taking all the income-related items into account, information on current earnings was successfully collected for 94.5% of the participants in PIAAC. Thus, no item NRBAs were required in Germany.

Basic Nonresponse Bias Analyses

The goal of the basic NRBAs was to identify the potential for nonresponse bias in the data set after data collection and to select those variables that could most effectively reduce the bias when used for nonresponse adjustment.

Preparation of Analyses: The Selection of Auxiliary Variables

As specified by the international Consortium, the auxiliary variables used in the nonresponse analyses needed to satisfy a specific set of criteria: (1) In order to analyze differences in response behavior, information has to be available for both respondents and nonrespondents. Usually, this information relates to variables that are already available from the sample frame, fieldwork process, or external sources, and which may be matched to each single case of the gross sample. (2) The auxiliary variables have to be related both to response behavior and to the central study outcome (in PIAAC: proficiency). This is a prerequisite to effectively reduce nonresponse bias. (3) Variables were to include no or only very little missing data. (4) Only variables with a high degree of accuracy were to be used for the analyses. (5) As a minimum requirement, the variables age, gender, education, employment, and region had to be included in the analyses. If these variables were not available for each individual case, it was possible to use an aggregated area level variable instead (OECD, 2010a).

46 0.5% are not part of the sampling frame because they are illegal immigrants and not included in official registers. The remaining 2% undercoverage is due to target persons who were initially part of the sample frame but moved during data collection and whose new address could not be traced.

In PIAAC Germany, three main sources of auxiliary variables were used:

1) *Registry information and information from the sampling frame*: Socio-demographic (age, gender, and citizenship) and geographic information (region and municipality size) were taken from these sources.
2) *Case folders*: The interviewers were required to complete a case folder for each target person, in which they primarily recorded the contact history (see Section 3.3). Additionally, interviewers were required to enter information about the target person, for example, type and condition of the dwelling, or the interviewer's subjective assessment of level of education and social class.
3) *Commercial database*: A consumer marketing database (Microm) provided economic, socio-demographic, and psychographic information about areas such as estimates of social class, or purchasing power (Microm MARKET & GEO, 2011).

Of these three data sources, registry variables have the highest quality. This information is collected by official sources, provided at the individual level, and has very few missing values. The case folder information, such as the assessed social class, is based on the interviewer's subjective judgment. Moreover, interviewers were advised to complete the case folder's questions on the target person's characteristics *prior* to the first contact. This requirement aimed at obtaining comparable information for respondents and nonrespondents. However, it also made assessments of the target persons' social class or educational status extremely difficult, because interviewers had to base their judgments solely on external factors, such as characteristics of the target person's dwelling and the neighborhood. Microm variables had two quality limitations. First, between five to approximately 500 households were combined (Microm MARKET & GEO, 2011). Second, data were not available for approximately 5% of the sampled units.

Because the auxiliary variables need to be significantly related both to the central study outcome, proficiency, and to response behavior, the first step of the analysis consisted of examining the relationship of the variables from these three sources with proficiency. Since the basic NRBAs had to be conducted at a time when the proficiency scores were not yet available, a proxy variable for proficiency had to be used. Analyses with German field test data revealed that the best approximation for proficiency was the highest general school leaving qualification in Germany ("education"): Its correlation with the field test proficiency measure was $r = .44$, ($p < .001$).

The majority of the auxiliary variables available proved to be significantly correlated with the highest general school leaving qualification ("education") at the bivariate level. Because several of the Microm variables were based on similar information with similar content, multicollinearity between these variables was investigated, and multinominal logistic regressions were conducted to identify the most important predictors of education. The variables with the largest predictive power constituted the initial set of auxiliary variables that were used as predictors of response behavior in the NRBAs described in the following section.

Analyses

After the identification of a suitable set of auxiliary variables related to the study outcome, the relationship of these variables to response behavior was examined. Both bivariate and multivariate analyses were conducted in order to determine the strongest predictor variables of participation in PIAAC. The analyses required by the international Consortium included (Mohadjer et al., 2013a):

1) chi square tests of auxiliary variables with response behavior,
2) response rate analyses for different subgroups,
3) classification tree analysis to identify subgroups with low response rates, and
4) logistic regression to model the relationship between response propensity and the auxiliary variables.

The bivariate analyses (1 & 2) were used to assess the relationship between each individual auxiliary variable and response propensity, and thus provided descriptive information on the bias present in the data set. In addition, interactions between the candidate variables were taken into account with the multivariate analyses (3 & 4). Thus, the multivariate analyses identified the strongest predictor variables of response behavior, i.e., the variables that reduced bias most effectively when used in weighting.

Results

The *chi square tests* (1) revealed that almost all auxiliary variables were significantly related to response behavior. However, among those variables required by the international Consortium for the analyses (age, gender, region, education, and unemployment),[47] only age, gender, and education proved to be significantly related to response behavior ($p < .05$), whereas region ($p = .50$) and unemployment ($p = .25$) showed no significant effect.

The chi-square test results were supported by the *analyses of response rates per subgroups* (2). These analyses also showed that participation in PIAAC Germany varied as a function of age, gender, citizenship, educational level, and municipality size. Among age groups, the 16 to 25 year-olds showed the highest (base-weighted) participation rate (67.9%), compared to the other age groups (51.8% – 54.2%). Regarding gender, the participation of women was slightly higher, compared to men (56.3% versus 54.2%). Furthermore, German citizens (56.5% versus "non-German citizens": 48.5%) and those with higher levels of educational attainment (60.9% versus "low": 50.7%) were more likely to participate in PIAAC than their counterparts. Regarding municipality size, the response rate was lowest in large cities with 500 000 and more inhabitants (49.3%) and highest in small municipalities with 2 000 – 4 999 inhabitants (60.3%).

In addition to these descriptive analyses, multivariate nonresponse analyses were conducted. Because these analyses aimed at identifying a set of variables useful for weighting, only registry and sampling frame variables—age, gender, citizenship, and municipality size—were investigated. Both the auxiliary information collected by interviewers as well as the Microm variables did not satisfy the high quality requirements for weighting information. As previously mentioned, this is due to the subjective nature of the interviewer's assessments, the varying levels of aggregation, and the issue of missing data in the Microm database.

In the *classification tree analysis* (3), response behavior was predicted by the independent variables mentioned above. In a classification tree predicting response behavior, the gross sample is split into a number of subgroups with differing response rates. In Germany, the first split criterion was age, followed by subgroups based on citizenship and municipality size. Gender did not have enough predictive power to appear as a split criterion in the classification tree analysis.

The *logistic regression model* (4) significantly predicted response behavior (Prob > $F = .000$). As Table 3.12 shows, all variables entered into the model had a significant effect on response propensity in PIAAC; however, gender ("male": $p = .02$) was only significant at the 5% level, whereas all other variables were significant at the 1% level.

47 "Education" was measured here by the interviewers' assessment of the target persons' level of education as given in the case folders. "Unemployment": Microm variable displaying the aggregated unemployment rate in a market cell.

Table 3.12. Regression Model of Response Behavior

Response behavior	Coefficient	Linearized *SE*	*t*	*p* value
Age	-.0137763	.0015765	-8.74	0.000
German	.4723318	.0796205	5.93	0.000
Male	-.1109863	.0478911	-2.32	0.021
Municipality size	-.0672663	.0157497	-4.27	0.000
_cons	.6260526	.1269310	4.93	0.000

Notes. Response behavior coded as 1 = response and 0 = nonresponse. *SE* = standard error. *t* = t-statistic. _cons = constant.

Based on the results of the multivariate analyses, age, citizenship, and municipality size were chosen as variables to be used for producing nonresponse weights. Gender was omitted as a variable for the nonresponse weighting step, because the number of weighting variables was to be kept low and this variable was the least important in predicting response behavior.

Extended Nonresponse Bias Analyses

Countries with a response rate lower than 70% were required to conduct extended NRBAs. These were performed after the weights had been calculated and they aimed at investigating the effectiveness of weighting adjustments on nonresponse bias reduction. As required by the international Consortium, the extended NRBAs included the following analyses (Mohadjer et al., 2013a):

1) comparison of estimates before and after weighting adjustments,
2) comparison of final weighted estimates with Microcensus data,
3) correlations of auxiliary variables and proficiency estimates,
4) comparison of estimates from alternative weighting adjustments,
5) analysis of variables collected during data collection,
6) level-of-effort analysis, and
7) calculation of the range of potential bias.

For the *comparison of estimates before and after each weighting step* (1), potential changes in the distributions of the weighting variables when different weights are used were examined and evaluated. Results are displayed in Table 3.13.

In order to identify initial nonresponse bias (similar to the basic NRBAs), the unknown eligibility-weighted distribution of variables used for weighting such as age or gender—for respondents—was first compared to the corresponding distribution in the eligible sample data (see columns (b) and (a) in Table 3.13). The results confirmed the findings of previous analyses: The initial respondent sample was slightly biased with regard to age, citizenship, and municipality size. In particular, the 16 to 25 year-olds (19.5% versus 16.1%) and German citizens (92.4% versus 91.1%) were slightly overrepresented in the respondent sample, compared to the eligible sample. On the other hand, the 46 to 55 year-olds (23.5% versus 25%), non-Germans (7.6% versus 8.9%) and persons living in large cities with at least 500 000 inhabitants (12.7% versus 14.4%) were somewhat underrepresented. Data about the highest general school leaving qualification were not available on the sampling frame.

In the next step, the nonresponse-weighted distribution of the respondent characteristics was compared to the gross eligible sample (see columns (c) and (a) in Table 3.13).

Table 3.13. Distribution of Sample Cases at Each Weighting Step and Microcensus Data, by Key Auxiliary Variables

Variable	Used in weighting	UEW: Eligible sample (a)		UEW: Respondents (b)		NRW: Respondents (c)		CAL: Respondents (d)		Micro-census 2010 (e)
		%	*SE*	%	*SE*	%	*SE*	%	*SE*	%
Age	UEW/ NRW/CAL									
16 to 25		16.1	0.3	19.5	0.5	16.2	0.5	17.3	---	17.3
26 to 35		18.9	0.4	18.2	0.6	18.8	0.6	18.0	---	18.1
36 to 45		22.0	0.4	21.6	0.6	22.3	0.6	23.0	---	23.1
46 to 55		25.0	0.4	23.5	0.6	25.0	0.6	23.7	---	23.6
56 to 65		18.0	0.4	17.2	0.5	17.8	0.6	17.9	---	17.9
Gender	CAL									
Male		49.9	0.5	48.6	0.7	48.4	0.7	50.4	---	50.4
Female		50.1	0.5	51.4	0.7	51.6	0.7	49.6	---	49.6
Citizenship	UEW/ NRW									
German		91.1	0.4	92.4	0.5	91.1	0.5	90.6	0.5	89.3
Not German		8.9	0.4	7.6	0.5	8.9	0.5	9.4	0.5	10.7
Municipality size (number of inhabitants)	UEW/ NRW									
1 – 1 999		6.0	1.1	6.0	1.1	6.1	1.1	6.3	1.4	5.8
2 000 – 4 999		9.4	1.3	10.3	1.5	9.4	1.4	9.5	1.6	9.0
5 000 – 19 999		26.4	2.1	27.8	2.2	26.3	2.1	26.8	2.7	25.7
20 000 – 49 999		18.9	2.0	19.0	2.1	19.0	2.1	18.8	2.1	18.4
50 000 – 99 999		9.3	1.5	9.3	1.5	9.3	1.5	9.2	1.7	8.9
100 000 – 499 999		15.6	1.5	14.9	1.4	15.6	1.5	15.5	2.1	15.5
500 000 – 99 999 999		14.4	1.2	12.7	1.1	14.3	1.2	13.9	1.7	16.8
Region[1]	CAL									
North		40.6	2.8	40.2	2.8	40.7	2.8	40.4	---	40.4
South		41.9	2.8	41.6	2.8	41.1	2.8	42.1	---	42.1
East		17.6	2.2	18.2	2.3	18.1	2.2	17.5	---	17.5
Education	CAL									
Low		n/a	n/a	26.5	0.9	26.5	0.9	32.4	---	32.5
Medium		n/a	n/a	35.4	1.0	35.5	1.0	34.0	---	33.8
High		n/a	n/a	35.3	0.8	35.6	0.8	30.4	---	30.4
Student (up to high school)		n/a	n/a	2.8	0.2	2.4	0.2	3.3	---	3.3

Notes. UEW = unknown eligibility weight. NRW = nonresponse weight. CAL = final weight. *SE* = standard error, computed taking the complex sample into account. --- = the estimate has no variation across replicate weights. n/a = not applicable.

[1] North = Hamburg, Schleswig-Holstein, Bremen, Lower Saxony, and Berlin (West); South = North Rhine-Westphalia, Hesse, Saarland, Rhineland-Palatinate, Baden-Württemberg, and Bavaria; East = Berlin (East), Mecklenburg-Western Pomerania, Brandenburg, Saxony, Saxony-Anhalt, and Thuringia.

Information provided by official registries on age, citizenship, and the frame variable municipality size was used for the nonresponse weighting. The comparison of the nonresponse-adjusted data with the eligible full sample shows that this adjustment effectively reduced the initial nonresponse bias in the estimates of the variables used for the adjustment. The adjusted estimators are virtually identical to the distributions in the eligible sample for the variables chosen for nonresponse weighting; the maximum difference is 0.3 percentage points.[48] Furthermore, only minor changes in the distributions of variables not used in nonresponse weighting were detectable.

In the *comparison of final weighted estimates with Microcensus data* (2), the data were weighted with the final weight and compared to the corresponding Microcensus data.[49] This made it possible to examine whether the final weighted PIAAC data matched official external data for Germany and were thus representative for the 16 to 65 year-olds in Germany. Results are displayed in columns (d) and (e) of Table 3.13.

The comparison showed that any potential undercoverage and nonresponse bias with respect to the variables used for the calibration (age, gender, region, and education) was effectively eliminated. In fact, the PIAAC data are almost perfectly matched to Microcensus data for these variables.[50]

The calibration also reduced the coverage bias in the eligible sample regarding the variables used in nonresponse weighting, i.e., citizenship and municipality size. For example, the data for the eligible sample shown in column (a) indicate that 91.1% of respondents were Germans, compared to 89.3% according to the population figures for 2010, as given in column (e). After the weighting steps, the PIAAC estimate (90.6%), as shown in column (d), was closer to the population distribution. Regarding municipality size, the eligible sample was already a good representation of population data and the final weighted data do not differ substantially from Microcensus data. Only the category of large cities with 500 000 inhabitants or more was somewhat underrepresented in the eligible sample (14.4% versus 16.8%). This underrepresentation could not be eliminated by weighting. The final weighted data set still showed slight deviations for large cities, compared to the Microcensus (13.9% versus 16.8%).

To summarize, these analyses demonstrated that the final PIAAC weights effectively eliminated potential undercoverage and nonresponse bias regarding the variables used in the calibration and reduced bias in variables used in previous weighting steps.

A further analysis consisted of the computation of *correlations of auxiliary variables and proficiency estimates* (3).

As mentioned above, in order to effectively reduce bias in PIAAC, it was necessary to identify and use auxiliary variables for weighting that were significantly correlated with proficiency. At the time the weighting variables were chosen, the final proficiency scores were not yet available and the best available proxy (see basic NRBAs), i.e., the highest gen-

48 These minimal differences are due to two reasons. First, weighting cells had to be collapsed across age and municipality size. Second, "respondents" and the "eligible sample" were coded differently during the adjustment of the eligible sample (in the nonliteracy-related weighting step) than in the extended NRBAs. In this analysis of the extended NRBAs, the literacy-related nonrespondents with age and gender successfully collected, as well as breakoffs during the assessment due to literacy-related reasons, were included as respondents, whereas they were completely excluded during the benchmarking process. The weighting checks (using identical definitions as applied during the benchmark) showed deviations of up to 0.2 percentage points due to collapsing.

49 The Microcensus is a mandatory representative survey among 1% of households in Germany. For the comparisons, data from 2010 were used.

50 Deviations of 0.1 to 0.2 percentage points are due to the fact that weighting cells had to be collapsed across categories of age groups, gender, and the level of education. Another reason for minor deviations is rounding, because Microcensus data provided estimates rounded to 1 000 persons.

eral school leaving qualification, was used instead. The final proficiency scores were, however, available when the extended NRBAs were conducted.

This analysis aimed at validating the relationship of the auxiliary variables used in weighting with proficiency. For pragmatic reasons, the participating countries were instructed by the international Consortium to use only the first *plausible value*[51] for literacy (in the following referred to as proficiency) in the subsequent analyses. Results showed that weighting variables were highly correlated with proficiency, with a correlation of 0.61.[52] This further supports the conclusion that the weighting strategy was appropriately chosen to reduce bias in proficiency.

Correlations between non-weighting variables, for example from the background questionnaire, and proficiency were also examined. A strong correlation of such variables with proficiency could indicate a residual potential for bias in the proficiency scores, after weighting adjustment, because they were not used in weighting. Of the different variables analyzed, the ISCED classification,[53] for example, showed a significant correlation with proficiency ($p < .001$). In this context, the correlation between a more differentiated "region" variable, based on Germany's 16 federal states, and proficiency was also computed. The coarsened "region" variable had not shown a significant relationship with proficiency. The variable based on the 16 federal states did, indeed, show a significant relationship with proficiency ($p < .01$). Based on these analyses, potential alternative weighting variables were identified for the following calculation of alternative weights.

In a fourth set of analyses, *estimates from alternative weighting adjustments* (4) were compared, in order to investigate whether alternative procedures would have achieved a greater reduction in bias. Therefore, the final weighted data were post-stratified using variables from analysis (3) that were not used in weighting but found to be highly correlated with the proficiency score, e.g., ISCED and the detailed "region" variable "Federal States". The ISCED categorization was of particular interest as an alternative weighting variable because, in PIAAC, international education comparisons are based on this variable, whereas a national education variable ("highest general school leaving qualification") was used for weighting of the German data.

The comparison of the mean proficiency score weighted by the final PIAAC weight with the mean proficiency score weighted with the alternative weights yielded no or only very small changes. The most important, but nevertheless small change, was caused by the alternative weighting using ISCED, which reduced the mean proficiency score by two score points from 270 to 268.[54] Poststratification using "Federal States" showed no change in the mean proficiency score.

In sum, the alternative weighting adjustments provided further indication that bias was already effectively reduced by the chosen weighting adjustment and an alternative weighting would not have yielded very different results with the set of variables selected for this analysis.

In the *analysis of variables collected during data collection* (5), the interviewers' subjective assessments of the nonrespondents' level of education and social class, as well as the information on type and condition of the nonrespondents' dwelling were compared to those of

51 Plausible values are estimated proficiency scores obtained by IRT modelling.

52 Correlations were computed as the square root of R-square resulting from a linear regression with proficiency as the dependent variable and the weighting cells used for nonresponse adjustment and for calibration as independent variables (see Mohadjer et al., 2013a, p. 17). Estimates used in calibration have no variation across replicate weights. Therefore, a significance test is not appropriate.

53 For more information on the ISCED (International Standard Classification of Education) see UNESCO Institute for Statistics (2006).

54 With a standard deviation of $SD = 48$ for the first plausible value of literacy, this change can be considered to be small.

respondents. This analysis explored potential reasons for nonresponse in more detail. The results could be useful for tailoring efforts to increase response rates in future cycles of PIAAC. The analysis revealed that, compared to respondents, nonrespondents on average have lower assessed levels of education and social class, and have a higher tendency to live in large (apartment) buildings. For example, 18.3% of nonrespondents lived in apartment buildings with nine or more units, compared to only 13.8% of respondents. The latter were more likely to live in single and terraced houses than nonrespondents (53.8% versus 45.7%).

However, it has to be kept in mind that some of these results must be treated with caution, because the data are based on interviewers' subjective assessments of target persons made before they had established an initial contact (see basic NRBAs).

A further analysis aimed at verifying whether the literacy-related cases had been assigned the correct disposition code. This served as an indication that the literacy-related nonresponse adjustment was conducted accurately. In this analysis, central characteristics of literacy-related nonrespondents were compared to other (nonliteracy-related) nonrespondents, to examine whether they belonged to socio-demographic groups in which language problems tend to be more common. This was confirmed for Germany. For example, the majority of literacy-related nonrespondents was not German (58.3%), and only 41.7% were Germans. Among the nonliteracy-related nonrespondents, only 9.5% were foreigners and 90.5% were Germans. Another example is that literacy-related nonrespondents were more often evaluated by the interviewers as belonging to a lower social class (77.7%) than nonliteracy-related nonrespondents (42.5%).

The correct assignment of the disposition codes showed that the literacy-related nonrespondents were treated appropriately during the separate literacy-related nonresponse adjustment and that nonresponse bias was thus reduced with this weighting step (Mohadjer et al., 2013a).

In the *level-of-effort analysis* (6), the mean proficiency of late and early respondents was compared. Target persons interviewed during the main working phases were defined as early respondents; target persons interviewed during the re-issue phases were defined as late respondents.

Differences in proficiency levels may indicate that fieldwork efforts to gain late respondent participation were effective in reducing nonresponse bias. At the same time, they indicate that some level of nonresponse bias might still be present in the data. However, this is only true to the extent that differences between early and late respondents reflect differences between respondents and nonrespondents. In order to test the assumption that late respondents are similar to nonrespondents, we also analyzed whether the late respondents shifted the realized sample of early respondents closer to known population totals with respect to key socio-demographic characteristics such as education, work status, and citizenship. In addition, differences between early and late respondents in the distributions of these variables were tested for significance.

Results shown in Table 3.14 reveal that the mean proficiency in the group of late respondents (high level-of-effort) was lower than in the group of early respondents (low level-of-effort), but the difference was not significant.[55] Similarly, differences between early and late respondents in the distributions of education, work status, and citizenship were all not significant. Analogously, adding the late respondents to the realized sample of the early respondents shifted the distribution of these variables only slightly closer to that of the Microcensus.

55 The difference of approximately 4 score points was only significant at the 10% level (p = .099).

Table 3.14. Differences Between Low-Level-of-Effort and High-Level-of-Effort Respondents in Proficiency

	Estimate of population proportion			Comparison data		Proficiency estimate		
	Proportion among low level-of-effort	Proportion among high level-of-effort	*p* value of the F statistic	All respondents	Microcensus 2010	Low level-of-effort	High level-of-effort	*p* value of the t-test
Mean	n/a	n/a	n/a	n/a	n/a	270.6	266.4	0.099
Education								
Low education	31.3	33.9	0.137	31.6	32.5	239.7	233.5	0.134
Medium education	34.8	31.7		34.4	33.8	269.2	263.6	0.047
High education	30.5	32.5		30.7	30.4	302.1	302.9	0.804
Student (up to high school)	3.5	2.0		3.3	3.3	285.4	274.3	0.458
Work status								
Employed	75.4	75.5	0.479	75.4	71.2	274.9	272.7	0.394
Unemployed	4.1	5.2		4.2	5.5	256.3	245.1	0.209
Out of labor force	20.5	19.3		20.4	23.3	257.5	247.4	0.107
Citizenship								
German	91.6	89.9	0.370	91.4	89.3	273.4	270.5	0.244
Not German	8.4	10.1		8.6	10.7	239.9	230.1	0.320

Notes. n/a = not applicable. All respondents: Initial literacy-related nonrespondents with age and gender collected (n = 86) were excluded from the group of respondents in this analysis.

Overall, these results indicate that the late fieldwork efforts to obtain late respondents ranged from neutral to somewhat successful in adding more cases to the realized sample and that nonresponse bias was thus potentially slightly reduced by late fieldwork efforts.

The *calculation of the range of potential bias* (7) provides an estimate of how much bias in proficiency, due to nonresponse and undercoverage, potentially remains in the data set after weighting adjustments (OECD, 2010a).

The analysis explored how much the mean proficiency score would deviate from the actual score if all nonrespondents and persons not included in the sampling frame (in the following called "nonparticipants") had taken part in PIAAC. Because the proficiency levels for nonparticipants are not known, the estimated bias was calculated as a range. This range was determined by assuming extreme proficiency outcomes for nonparticipants. For the lower bound, it was conjectured that this group would have achieved a result corresponding to the 10th percentile of the realized sample's mean proficiency score. For the upper bound, the 90th percentile was taken as the reference.

The calculations were computed for each weighting cell, taking the response rates of each weighting cell into account. The more homogenous the weighting classes regarding proficiency (i.e., the better the weighting classes are defined) and the higher the response rates, the smaller the resulting range of bias (Mohadjer et al., 2013a).

The estimated range of bias in proficiency, after weighting adjustments, was 53 points for Germany. The range analysis suggests that if all nonparticipants had reached a literacy proficiency level corresponding to the 10th percentile of the realized sample, the weighted mean proficiency estimate for literacy would have been 243, i.e., 27 points lower than the actual score of 270. If all nonparticipants had reached a proficiency level corresponding to the 90th percentile of the actual PIAAC respondents, the proficiency estimate would have been 296, i.e., 26 points higher.

Given the extreme assumptions about the nonparticipants' score, and the moderately "low" response rate of 55%, this range can be considered to be relatively small. It indicates that weighting cells were well defined and that only a low potential for bias in proficiency remains after weighting (OECD, 2013c, Appendix 7).

The international Consortium additionally calculated the range of bias *before* weighting adjustments for countries that conducted extended NRBAs. The range of bias for Germany before weighting was estimated to be 121 points (Mohadjer et al., 2013a). The strong reduction of the range through weighting is a further indication that weighting adjustments were successful.

In sum, the analyses conducted in the extended NRBAs suggest that the adjustments performed with the adopted weighting strategy effectively reduced initial nonresponse and undercoverage bias in the German PIAAC data set, allowing for inferences to the target population of 16 to 65 year-olds in Germany.

3.4.3 Design Effect and Effective Sample Size

Besides low bias, a low variability of survey estimates is an indicator that sample data are of high quality. The variability of estimates largely depends on the sample size, with smaller samples yielding higher variability in the estimates (Mohadjer et al., 2013a). Given equal sample sizes, however, the variability of survey estimates from complex sample designs is larger than that of estimates from simple random samples. This is due to the clustering that occurs in complex sample designs, because the sampled units in a cluster are generally more homogeneous with regard to certain characteristics than sampled units selected by simple random sampling. More homogeneity among selected elements in the cluster decreases the precision of the estimate, i.e., increases its variance (Harter, Eckman, English, & O'Muircheartaigh, 2010).

The effect of the sampling design on the variance of an estimator is called the design effect. Equation (3) shows that the design effect is the ratio of the variance of the estimate obtained from the complex sample to the variance of the estimate from a simple random sample, assuming identical sample sizes (Mohadjer et al., 2013a, p. 29). For a statistic *t,* the design effect is calculated as

$$Deff(t) = \frac{Var_{Complex}(t)}{Var_{SRS}(t)} \quad (3)$$

The reduction of precision due to complex sampling can also be expressed by the effective sample size. This is the sample size for which a simple random sample would yield the same sampling variance as a complex sample (Mohadjer et al., 2013a, p. 29):

$$Effn(t) = \frac{n}{Deff(t)} \quad (4)$$

The international Consortium calculated the design effect for the three domains literacy, numeracy, and PS-TRE as well as the effective sample size for literacy for all countries participating in PIAAC.[56] The overall design effect includes the design effects due to sampling variance (i.e., unequal weights, stratification, and clustering) and imputation variance (OECD, 2013c, Appendix 7). Results for Germany show a design effect of *Deff* = 2.01 for literacy, *Deff* = 1.89 for numeracy, and a slightly higher design effect of *Deff* = 2.58 for

56 The variance of the complex sample design was calculated using the replicate weights. The simple random sampling variance was calculated for each of the 10 plausible values and then averaged, in order to obtain the overall simple random sampling variance.

PS-TRE; the effective sample size for literacy is n = 2 680, approximately half the size of the actual sample size n = 5 465 (Mohadjer et al., 2013a, p. 30).[57] These results are well in line with those from the other participating countries.

3.4.4 Variance Estimation

For the estimation of variance in PIAAC, two specifics have to be considered. First, the majority of countries used complex sample designs that introduced sampling variance, due to stratification or clustering. In PIAAC, the replication approach is used for estimating sampling variance. Second, proficiency is not captured in one single score that represents a proportion of correct answers in the assessment. Rather, proficiency was modelled based on the Item Response Theory, in which proficiency is seen as a latent trait, and several plausible values were imputed (See Section 4.3.1).

The combination of replication and multiple imputation approaches captures (a) variance due to the complex sample design, (b) differential selection probabilities, (c) the weighting adjustments, and (d) the measurement error through the multiple imputation of plausible values (Mohadjer et al., 2013c, p. 26).

Variance Estimation in Complex Samples: The Replication Approach

Two techniques are commonly used for variance estimation in complex survey designs: the Taylor Series Linearization and replication approaches. In PIAAC, the replication approach was used.

Several methods can be used to create replicate weights. These include jackknife methods, balanced repeated replication (BRR), or Fay's method. The choice of method depends on features of the sample design, such as whether stratification was used and how many PSUs there are per stratum (Heeringa, West, & Berglund, 2010). In Germany, the delete-one jackknife (JK-1) was used because the number of PSUs in each stratum varied between 0 and 2 (Mohadjer et al., 2013c).

For replication, the full sample was split up into replicate subsamples, based on sample design information. The number of replicates created depended on the sample design and was generally 60 or 80. In PIAAC Germany, 80 replicates were created, as was the case for most participating countries (Mohadjer et al., 2013c). The replicates for Germany were created by the international Consortium. After the creation of replicate base weights, all weighting adjustments that were conducted for the full sample were also conducted for each individual replicate sample (OECD, 2011c).

In order to compute the variance of an estimate, the sum of squared deviations between each replicate subsample estimate and the full sample estimate is calculated. Equation (5) specifies how the variance of any statistic of interest is estimated for JK-1 replication (adapted from Mohadjer et al., 2013c, p. 26):

$$Var_{samp}\left(\hat{\theta}\right) = \frac{r-1}{r}\sum_{i=1}^{r}\left(\hat{\theta}_i - \hat{\theta}_0\right)^2 \quad (5)$$

with:

r = number of replicates

$\hat{\theta}_0$ = full sample estimate

$\hat{\theta}_i$ = estimate for replicate i

57 The effective sample size was computed and provided by the international Consortium, using the design effect of the literacy domain, and was calculated as the number of cases with plausible values divided by the overall design effect for literacy (Mohadjer et al., 2013a).

Variance estimation with plausible values

For variance estimation involving proficiency scores, both the sampling and the imputation variance have to be taken into account.

As outlined in Mohadjer et al. (2013c), the estimator of the population mean is calculated as the average of the M plausible value means, as shown in equation (6). Equations (6) to (11) are taken from Mohadjer et al. (2013c, p. 29):

$$\hat{\bar{Y}}^* = \sum_{m=1}^{M} \hat{\bar{Y}}_m \Big/ M \quad (6)$$

The equation reflecting total variance for the mean estimate $\hat{\bar{Y}}_m^*$ is:

$$v\left(\hat{\bar{Y}}^*\right) = U^* + B\left(1 + \frac{1}{M}\right) \quad (7)$$

where U^* is the average of the sampling variances for each of the M plausible values:

$$U^* = \left(\sum_{m=1}^{M} U_m\right) \Big/ M \quad (8)$$

with U_m being the sampling variance of the estimated mean $\hat{\bar{Y}}_m$ for plausible value m ("within variance"). The "between variance" is:

$$B = \left[\sum_{m=1}^{M} \left(\hat{\bar{Y}}_m - \hat{\bar{Y}}^*\right)^2\right] \Big/ (M-1) \quad (9)$$

where the mean of each of the M PVs, $y_{l1}, y_{l2}, \ldots, y_{lM}$ for sample unit l is computed as

$$\hat{\bar{Y}}_m = \sum_{l \in s} w_l y_{lm} \Big/ \sum_{l \in s} w_l ; m = 1, \cdots, M \quad (10)$$

where s denotes the set of sample units, and m the plausible value. The subscript w_l is the weight for person l, and y_{lm} is the value of the mth plausible value for person l. The standard error is computed as the square root of the total variance:

$$\sqrt{v\left(\hat{\bar{Y}}^*\right)} \quad (11)$$

4. Data Management Processes, Data Products, and Overall Data Quality

After data collection, a well-structured data management process was put in place to accommodate the complex design of PIAAC. At the national level, interview data were checked for consistency, cleaned, and, where applicable, pooled with additional data from other sources (e.g., the case management system, coding, scoring) to produce the national database. At the international level, the national databases were cleaned, weighted,[1] scaled, and further processed. Additional variables were computed to simplify, but also to enhance the analytical options, and data confidentiality edits were implemented to protect respondents' identity. Finally, Public Use Files were produced for dissemination and made accessible to users by the OECD, together with PIAAC-specific data analysis tools to facilitate the analyses of data with up to 80 replicate weights and 10 proficiency scores per skill domain. Lastly, the overall data quality was evaluated in an adjudication process, in which the international Consortium, the Technical Advisory Board, the OECD, and the Board of Participating Countries were involved. This data adjudication revealed the overall *fitness-for-use* of the PIAAC data.

4.1 Data Management at the National Level

Each country appointed a National Data Manager who had the overall responsibility for the national data management processes. These included the cleaning, production, submission, and final check of the national database prior to dissemination. To guide national data management processes, the international Consortium provided a comprehensive data management manual, outlining the main scope of the required data management tasks, and it also offered a two-day in-person training session for further consolidation.

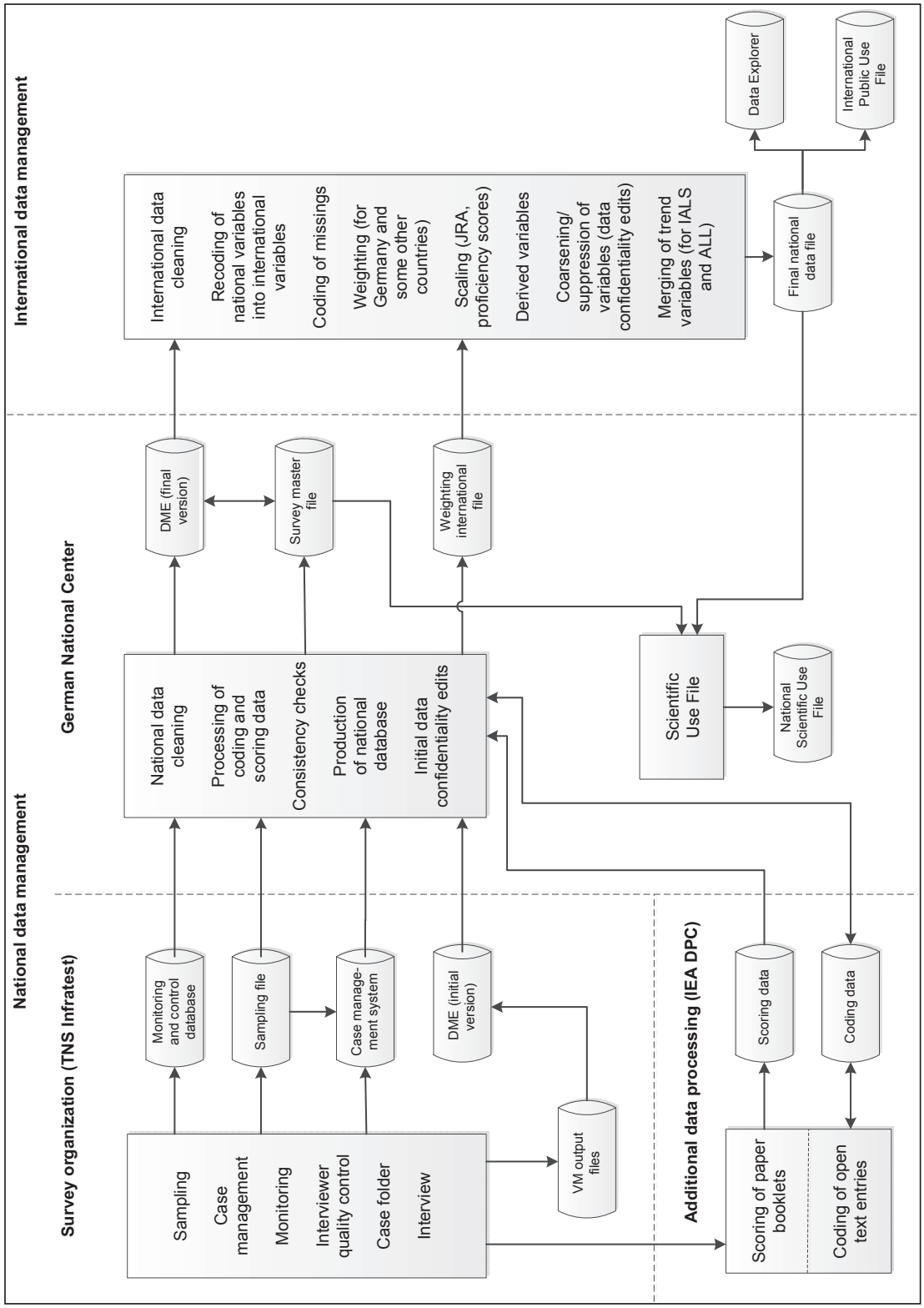
In Germany, the survey organization handled the initial data management tasks. A significant amount of data management was carried out by the German National Center. Data originated from a variety of data sources, for example, the case management system or the international interview software.

Data Management by the Survey Organization

As shown on the upper left-hand side of Figure 4.1, the four main databases handled by the survey organization were: (1) a monitoring and control database, (2) a sampling file, (3) a case management system database, and (4) an initial version of a customized relational database, the Data Management Expert,[2] which was made available to countries by the international Consortium. The first three databases were set up and maintained by the survey

1 The international Consortium conducted weighting only for certain countries. The other countries chose to do their own weighting.

2 For more information and details on the functionalities of the Data Management Expert, see Carstens and Daniel (2013).



Notes. ALL = Adult Literacy and Life Skills Survey. DME = Data Management Expert. IALS = International Adult Literacy Survey. JRA = Job Requirement Approach. VM = virtual machine.

Figure 4.1. Overview of data sources, data management activities, and data files

organization, and fed with data from sampling, case management, and information from the field, such as results of interviewer quality control or the case folder (entered via the case management system during fieldwork).

Interview data from the computer-assisted administration of the background questionnaire and from the computer-based assessment were collected in the virtual machine. The survey organization imported the virtual machine output files into the Data Management Expert database on a regular basis throughout the data collection period and provided it to the German National Center for further processing. Responses to the paper-based items were collected in booklets, which were pooled by home office. Booklets were then sent to IEA DPC, which was contracted by the German National Center to score the paper-based items (see Section 4.2). Scored data were subsequently submitted to the German National Center.

Data Management Tasks at the German National Center

Some data management tasks were undertaken at the German National Center before or during data collection. The first step consisted of adapting the customized relational database, in order to obtain a national Data Management Expert database that reflected all national adaptations and extensions implemented in the German instruments. The national codebook was then created on the basis of this national variable structure. The Data Management Expert software imported the complex virtual machine output files, extracted the data from various XML files into several tables, and stored these data. At the same time, databases maintained by TNS Infratest were updated regularly and made available to the German National Center for monitoring purposes and for fieldwork reports to the international Consortium.

After data collection had been completed, central data management activities were carried out to produce the national database (see middle section of Figure 4.1). First and foremost, data from the different sources were pooled (in a survey master file and the Data Management Expert database), checked for consistency, and cleaned. National data cleaning and consistency checks focused on data obtained outside of the virtual machine, because virtual machine data were checked and cleaned primarily by the international Consortium (see Section 4.3); only basic consistency checks were performed for data from the virtual machine.

The German National Center and the survey organization examined and resolved data inconsistencies between paper case folder information and electronic case management data during several reconciliation sessions. Reconciliation was also necessary to assign a final disposition code for some cases that had been re-worked in one of the re-issue phases. For example, if a case was a refusal during the main working phase and a non-contact during the re-issue phase, the final disposition "refusal" was assigned.

Several additional variables were created and imported into the Data Management Expert database. Some of these variables were computed by the German National Center; others, such as the scoring data, were generated by the sub-contractor and then further processed by the German National Center. IEA DPC was contracted by the German National Center to code open text entries of several background questionnaire variables into prescribed schemes or classifications, such as ISCO-2008 for occupation (see Section 4.2). The German National Center extracted the variables needed for coding from the Data Management Expert database, and prepared a file that was provided to IEA DPC. Coded data were then returned, checked for consistency, further processed, and added to the Data Management Expert database.

Variables computed by the German National Center included:

- sample design variables (e.g., identification numbers for PSUs, stratification variables, selection probabilities, and theoretical base weights);
- several flag variables: (a) quality control flags to earmark those cases that were selected for validation, also indicating the results of that validation, (b) sample flags to earmark whether cases were released in main working phase 1 or 2, (c) registry and exclusion flags to earmark cases with address-related problems and their final eligibility and exclusion status; and
- variables used for weighting (see Section 3.4).

Two central actions were taken to ensure adherence to national data confidentiality regulations prior to the submission of the national data file to the international Consortium. On the one hand, limited use and confidentiality affidavits were signed by each institution within the international Consortium that needed access to the German PIAAC database for international processing. The OECD also signed this affidavit. On the other hand, initial confidentiality measures were implemented through (a) suppression of direct identifiers such as names and addresses, (b) suppression or coarsening of variables that would allow the identification of respondents with minimal effort (e.g., fine-grained regional information), and (c) suppression of open text entries. However, most variables were left at the highest possible level of disaggregation at this stage, because this was crucial for the subsequent data processing steps.

After completion of all national data management tasks, a final Data Management Expert file was produced. This file underwent final validation checks and was then submitted to the international Consortium. At the same time, a final weighting file that contained information relevant for the weighting process (see Section 3.4), such as benchmarking data, was produced and delivered to the international Consortium. This submission included a detailed documentation of all national data management issues that were of relevance for the further international processing of the data (e.g., data cleaning and recoding activities, information on technical problems, suppression of variables).

The right-hand side of Figure 4.1 shows data management activities that took place at the international level after the submission of the national databases and which are described in more detail in Section 4.3.

4.2 Coding and Scoring

In Germany, most of the coding and all of the scoring activities were carried out by IEA DPC, as part of their sub-contract with the German National Center. We will now describe the national coding and scoring processes in more detail.

4.2.1 Coding of Education, Occupation, and Other Variables

Questions in the background questionnaire were administered in different formats: the majority with closed answer categories, and some with open text entries. As a part of the national data management tasks, open answer questions had to be converted into numerical information, and nationally adapted questions were recoded into the international answer scheme.

The international standards and guidelines (OECD, 2010b) specified how coding had to be carried out, to ensure standardized coding procedures in all countries. In order to guarantee an internationally comparable coding process, the international Consortium provided

two one-day in-person trainings on coding for National Centers and national coding supervisors. These trainings reviewed the coding specifications and recommended procedures for PIAAC, and focused, in particular, on the details of coding related to education, occupation, and industry, because these are some of the most important variables in PIAAC.

Codes produced by IEA DPC were subsequently checked for consistency by the German National Center. Additionally, a few variables such as regional information and derived variables were coded directly by the German National Center.

Coding of Educational Qualifications

For PIAAC, it was crucial to obtain an internationally comparable, reliable, and valid measurement of education. Questions on educational qualifications were posed in a variety of contexts in the PIAAC background questionnaire: (1) the highest educational level achieved, (2) educational qualifications currently pursued, (3) educational qualifications attempted but not completed, (4) educational qualifications in the last twelve months, (5) educational qualifications of respondents' mother and father, and (6) subjective indication of educational qualification normally required to obtain the job currently held by the respondent. As described in Section 2.2, questions on educational qualifications were administered according to the national education system. As part of the coding process, these national educational qualifications were mapped into an international classification, the *International Standard Classification of Education 97* (*ISCED-97*; OECD, 1999; UNESCO Institute for Statistics, 2006),[3] so that the educational qualifications were comparable across countries.[4] For Germany, this mapping was defined on the basis of information from the OECD (1999), from previous national studies (such as the German Microcensus), and in consultation with national education experts. These mapping definitions were converted into a recoding syntax by the international Consortium. Prior to the computation of the ISCED-97 codes by the international Consortium, this syntax was approved by the German National Center.[5]

Coding of Occupation

Occupation information had to be coded into the four-digit *International Standard Classification of Occupations 2008* (*ISCO-08*; International Labour Office, 2012).[6] This included occupation coding for the respondent's current or last job, and in Germany, also for the national extension on the occupation of the respondent's parents.

In Germany, a total of ten coders was trained for ISCO coding during a four-day in-person national training run by IEA DPC, with a subsequent extensive practice phase to consolidate their knowledge. Each coder was required to work through approximately 5 000 coding assignments for training purposes before carrying out the actual coding for PIAAC. Items that were not coded reliably were discussed and underwent a further practice cycle.

The subsequent procedure for coding the PIAAC occupation data was as follows:

1) All data were manually coded, i.e., there was no automated coding of answers with a list of conventions.
2) The coding was conducted strictly item-wise.
3) The coding sequence of answers per item was randomized to ensure that a doubtful, contingent decision did not affect several cases at once and, thus, systematically.

---

3 In 2011, a new version of ISCED, ISCED 2011, became available. However, it was not available when the preparations for PIAAC started.

4 The ISCED variables for the different educational questions (listed above) in the international master background questionnaire are: (1) B_Q01a, (2) B_Q02b, (3) B_Q03b, (4) B_Q05a, (5) J_Q06b and J_Q07b, and (6) D_Q12a.

5 A missing code was assigned for cases in which insufficient information was provided by the respondent.

6 Responses were coded directly into the international classification ISCO-08 (i.e., no interim national classification and crosswalk was used).

4) In order to avoid a systematic distortion of the coding results, simple *case of doubt rules* (e.g., "in case of doubt assign the numerically higher code") were not employed.
5) Coders were provided with some additional information (such as the industry or the economic sector of the respondent's job) to support them during the coding process.
6) In cases for which it was not possible to code the information into the four-digit ISCO, only fewer digits, and thus less detail, were provided. However, in comparison with many other studies, the information was not "filled up" with zeroes (trailing zeroes). This means that if it was not possible to distinguish between code 1311 and 1312, the code 131 and not 1310 was assigned, as the latter is not part of the ISCO scheme.

The international standards and guidelines (OECD, 2010b) specified that each country had to double-code[7] 50% of all the cases. However, in Germany, 100% of the information on occupation was double-coded, thus enhancing the quality of the coding process. For cases of disagreement, a supervisor at IEA DPC reviewed and resolved the discrepancies. According to the international standards, it was also necessary to calculate the interrater reliability. The requirement was that the interrater reliability be at least 90%. However, since Germany implemented full double-coding with reconciliation for all cases, this requirement did not apply.[8] The international Consortium reviewed and accepted the procedures and quality of the coding for ISCO in Germany.

As a standard quality assurance measure, the data were checked for "wild codes" (i.e., codes that do not exist in the ISCO scheme). An additional quality check compared the PIAAC ISCO distributions to those of the German Microcensus 2011. Both unweighted and base-weighted distributions of occupational groups were compared at the two-digit level. The international Consortium reviewed the German results and confirmed that the distribution was adequate.

Coding of Industry

Information on industry was coded into the four-digit *International Standard Industrial Classification of All Economic Activities* (*ISIC*), Revision 4 (United Nations Statistics Division, 2013a) as required by the PIAAC standards. An approach similar to that for ISCO coding was implemented. The coding was carried out by two independent coders who double-coded 100% of the cases. The first three coding procedures mentioned above for ISCO also applied to ISIC coding.

A total of three coders was trained for this task by IEA DPC. The national in-person training consisted of a one-day session, with an extensive two-week practice phase. During this period, ISIC coders processed approximately 1 000 answers for training purposes. As for ISCO, the reliabilities of the training cases were reviewed and, if necessary, there was an additional training cycle.

During the actual ISIC coding process, disagreements between coders were inspected and resolved by a supervisor at IEA DPC.[9] The German coding procedures and results for ISIC were also reviewed and accepted by the international Consortium.

Coding of Languages

In the PIAAC background questionnaire, there were several questions relating to languages that had to be coded into the international PIAAC coding scheme: (1) the first language(s) learnt and still understood, (2) the language(s) most often spoken at home, and (3) the language(s) spoken at home at age 16 (German extension). As mentioned in Section 2.2, the questions about the most common languages spoken in Germany were asked in a closed

7 Double coding means that two independent coders coded the information into ISCO.
8 The intercoder reliability before reconciliation was nevertheless calculated: It was 74% for the current occupation and 75% for the last occupation at the lowest common digit.
9 The interrater reliability before reconciliation was calculated and was 80% for the current job and 79% for the last job at the lowest common digit.

format in the background questionnaire. If the language mentioned by the respondent was not part of this list, it was captured as a free text entry. Both the list of languages and the answers to the open questions were mapped or coded into the three-digit alphanumeric code for languages, ISO 639-2/T alpha3, terminologic (Library of Congress Network Development and MARC Standards Office, 2013).

Coding of Countries and Other Regional Information

Several questions in the background questionnaire referred to country information: (1) country in which the highest foreign educational qualification was achieved, (2) country of birth, (3) parents' country of birth, and (4) citizenship, captured as country name (national extension). Country information was obtained either by using a country list or as an open entry response. All responses referring to country information were coded in an internationally comparable way, i.e., they were converted into the UN M49 scheme (United Nations Statistics Division, 2013b). The country names reflected the current name of the country, not the name of the country in the past (regardless of whether the question referred to the past, e.g., country of birth).

Regional information derived from sampling information was also coded: The respondents' place of residence was coded according to the OECD classification of geographical region, level TL2 (OECD, 2013d). In Germany, the TL2 scheme corresponds to the federal states.

Coding of Further Open Entry Information

In addition to the open entry questions mentioned above, there were a number of other open questions in the background questionnaire that required coding. These open questions were only asked if the respondent could not give an answer to a closed list. Most of these questions related to educational qualifications; two were job-related. For example, if respondents could not find their educational qualification in the list presented to them, or if they had a foreign qualification, they were asked to give their qualification as an open response. This information was then reviewed and, if possible, classified into the closed list.

National Adaptations

As described in Section 2.2, the German background questionnaire was adapted to fit the national context. For some questions, this meant that the number of response categories was adapted or that the answer mode was changed from a closed to an open format. One significant adaptation for the German background questionnaire was that respondents who answered that they had a vocational qualification (in one of the education questions) received an open and not a closed question about their field of education. IEA DPC recoded this information into the closed list according to *ISCED 1997 – Broad Fields of Education and Training* (Andersson & Olsson, 1999).[10] As an additional step, respondents who were currently in general education, and thus had not received this question, were automatically assigned the code "general programs" in the subsequent data management process at the National Center.

Derived Variables

The international Consortium and the OECD produced a number of derived variables for analytical purposes, for which national input was required. There were two variables in particular that required coding information from each country. The first captured the orientation of the qualification completed by the respondent, i.e., whether the qualification was vocational or general/academic (variable name: VET); the derivation of this variable was based on the German qualifications (for qualifications equal to ISCED 3 or 4). The second

10 The variables for fields of education (as given in the international master background questionnaire) are B_Q01b, B_Q02c, and B_Q05b.

captured hypothetical years of education (variable name: YRSQUAL). This variable was created on the basis of the average or the usual time that it takes to complete a qualification. In Germany, this definition was obtained in consultation with national education experts as well as staff from the Federal Statistical Office.

4.2.2 Scoring of Paper Booklets

As described in Section 2.1, the paper-based assessment was administered to respondents who did not have sufficient computer skills to take the computer-based assessment, or who were not willing to do the assessment on the computer. These respondents answered the items by writing their responses directly in the paper booklets. The handwritten responses had to be scored, which means that a value was allocated to each response. Contrary to the automatic scoring process implemented for the computer-based items (see Section 2.3.3), all paper booklets were scored by professional human scorers after the data collection phase (including the core booklet, which had been previously scored by interviewers during the course of the assessment, to determine routing in the paper branch). Thus, for each paper-based assessment case, in general, a total of three booklets was scored: a core booklet, a literacy or numeracy booklet, and a reading components booklet.

International Scoring Requirements

To ensure accurate and internationally comparable scoring results, a set of international standards and guidelines specified how the scoring process had to be implemented in each country (OECD, 2010b). Comprehensive information on how to score each paper item was provided in *item-by-item scoring guides* (Ferrari et al., 2013; Lennon & Tamassia, 2013; Tamassia, Lennon, & Yamamoto, 2013). These guides contained detailed information about the correct responses as well as some examples of both correct and incorrect responses. In each country, scoring guides were translated and adapted; it was important that scoring rules reflected national adaptations of items, such as changes in currencies or physical units. The scoring rules are an integral part of an item and are crucial to item functioning. All national scoring rules were verified by the international Consortium (see Section 2.4). The international Consortium also provided an in-person scorer training for countries and their lead scoring staff, introducing the scoring procedures for the paper instruments and reviewing the scoring rules and specifics for each paper item. Additionally, a scoring distribution list was set up, so that countries could consult the international Consortium during the scoring process, and which ensured that all countries received updates about new scoring issues and decisions.

The scores used for the paper instruments were identical to the scores for the computer instruments: score *1* was given for a *correct response,* score *7* for an *incorrect response,* and score *0* for an *omitted question*.

The general scoring rules can be summarized as follows: (a) correct responses had to be identical or equivalent to the information in the scoring guides, irrespective of spelling or grammar mistakes; (b) any response that was not equivalent to the correct response in the scoring guide was to be scored as incorrect; (c) if the respondents did not answer the item but made any kind of mark on the page, such as a dash or a question mark, the response was to be scored as incorrect; (d) if the page was left totally blank, the item was scored as omitted; (e) any supplementary information to the response needed to be judged by scorers on an individual basis, e.g., an otherwise correct response was disqualified if incorrect information was additionally included.

It was important to ensure scoring comparability both within and across countries. The international Consortium developed several possible scoring designs for countries' scoring process of the paper booklets. To establish within-country reliabilities, the international

standards and guidelines specified that a minimum of 600 sets of booklets be double-scored within each country (OECD, 2010b). Furthermore, prior to the scoring of the paper booklets, the international Consortium conducted reliability studies to ensure that scoring was comparable across countries (Tamassia et al., 2013). In these reliability studies, 60 so-called *anchor booklets*—consisting of 60 core booklets, 60 literacy booklets, and 60 numeracy booklets, all in English—were scored by bilingual scorers in each country. Countries submitted their scoring of the anchor booklets to the international Consortium. The Consortium evaluated the degree of scoring consistency between the participating countries, and concluded that the scoring was accurate and comparable across the countries (Tamassia et al., 2013).

Scoring of Paper Booklets in Germany

Scoring of the German paper booklets was carried out by IEA DPC. In Germany, double-scoring was carried out for *all* paper booklets, and thus exceeded the minimum of 600 double-scored booklets required by the international standards and guidelines. Two lead scorers who attended the international in-person scorer trainer meeting scored the anchor booklets for the international reliability studies mentioned above (both lead scorers were bilingual). They were also responsible for conducting national scoring training and supervision.

IEA DPC developed a six-scorer design (the international designs foresaw only two or three scorers), which was more appropriate for its standard processes. This design was approved by the international Consortium. For scoring, all booklets were organized in bundles (booklet sets consisting of core booklets and literacy booklets, or core booklets and numeracy booklets), and each bundle was double-scored. All scorers were previously trained for PIAAC. The national training was based on original items from the field test, and scorers were trained until scores for all items reached an acceptable scoring quality.

The scoring process in Germany started after data collection and took approximately one month. Paper booklets were sent directly by the survey organization via registered mail to ensure the highest possible level of security, and were then scanned and imported into the IEA DPC scoring software. Scorers entered scores electronically into the scoring software, and scoring reliability of the actual scoring was checked on a weekly basis; all scoring discrepancies were resolved by the lead scorers.[11]

The German National Center reviewed the scoring data delivered by IEA DPC and carried out some spot checks of the scoring and reconciliation results. The data were then captured in the Data Management Expert database. Scoring data, scoring resolution reports of all score discrepancies, and reliability results were subsequently submitted to the international Consortium.

Coding of the Reading Components

The reading components booklets consisted exclusively of closed format questions. Thus, no genuine scoring process was required. Each (multiple choice) item was straightforwardly coded as correct, incorrect, or missing. After scanning the booklets, responses for the reading components were double-captured in the IEA DPC software and automatically recoded. These data were also integrated in the Data Management Expert database.

4.3 Data Processing After National Database Submission

As mentioned previously, after all national data management activities had been finalized, each country submitted its database to the international Consortium. The Consortium was responsible for the comprehensive international data management activities, including deriving the final international databases from the national ones.

11 Scoring resolution was allowed because all booklets were double-scored.

The international Consortium first carried out a comprehensive set of data checks. All inconsistencies were resolved in cooperation with countries; there were only very few, minor issues in Germany. The international Consortium recoded some national variables into international variables according to national specifications (e.g., ISCED classification). Furthermore, internationally comparable missing schemes were assigned to background questionnaire variables. Whereas missing codes such as "don't know" or "refused" were directly assigned by interviewers during the interview, some data were missing by design for specific variables. Due to the routing in the background questionnaire, some participants did not respond to particular background questionnaire items; for such cases, the additional missing code "valid skip" was assigned. A similar missing scheme was also applied for the assessment variables. It differentiated between missing by design, omitted responses (responses to computer-based items with a respondent interaction time of less than five seconds were also treated as omitted responses), not reached, or not attempted (Yamamoto et al., 2013c).

For several countries, including Germany, the international Consortium conducted the weighting and computation of replicate weights (see Section 3.4). Scaling of both the cognitive items and central background questionnaire scales, specifically the Job Requirement Approach items, was also carried out by the international Consortium. Furthermore, both the international Consortium and the OECD computed a number of derived variables for analytical purposes.[12] Finally, trend variables were created to allow, in principle, for direct comparisons of specific variables with the previous surveys of adult skills, IALS and ALL (for information on trend variables see OECD, 2013c, Appendix 4).[13]

After finalizing these international data processing activities, each country received a preliminary final version of their national data file and could request the coarsening and suppression of variables for reasons of data confidentiality. The international Consortium subsequently produced fully labeled integrated national data files with an international database structure in SPSS and SAS format in which these confidentiality edits were accordingly implemented. This national database was checked and verified by the National Center in several iterations and was the basis for the dissemination database.

4.3.1 Proficiency Scaling and Proficiency Scores in Brief

After countries submitted their national databases to the international Consortium, and after data cleaning and weighting, the Consortium established separate proficiency scales for each assessment domain using Item Response Theory (IRT) scaling and latent regression models (population modeling), producing a set of plausible values for literacy, numeracy, and PS-TRE. The assessment results for each domain are reported along the corresponding proficiency scale. The proficiency scale is regarded as an ability continuum, and is differentiated into several proficiency levels. These levels are useful for interpreting the proficiency results and are described in terms of the type of tasks adults with proficiency scores within a defined range are likely to complete successfully.

For the reading components, IRT procedures were not applied. Instead, two results were produced, one indicating the percentage of correct responses, the other the overall timing of the reading components assessment.

12 A condensed summary of variables released in the Public Use File, including the derived variables, is given in Table 23-2 of the international technical report (Carstens, Daniel, & Gonzalez, 2013, pp. 6-7).

13 Zabal et al. (2013) specified a number of methodological constraints regarding the German IALS data, indicating that more research is needed to evaluate the impact of these issues on possible trend analyses based on these data.

Scaling

As mentioned above, scaling of the cognitive data and related procedures were undertaken at an international level, and it is beyond the scope of this report to provide details on these activities. Only a very succinct summary of central scaling procedures and related issues will now be given, on the basis of the international technical report (Yamamoto et al., 2013a, 2013b, 2013c) and the Reader's Companion (OECD, 2013b). We refer the reader to these documents for detailed accounts of these activities.

The PIAAC assessment design was based on a variant of matrix sampling. This design included multistage adaptive testing for a part of the cognitive assessment, two different modes, and items linking back to IALS and ALL. Within this design, respondents only worked on a subset of items, with varying difficulties, from the entire pool of items, thus reducing testing time and respondent burden, while at the same time ensuring broad construct coverage through a large pool of items. The matrix-sampling design allows a more efficient estimation of proficiencies when *populations* are analyzed. However, it does not yield a reliable and exact measure for *individual* performance. This is an important distinction to keep in mind.

Due to the nature of the matrix design, respondents' performance cannot be simply determined by the number of correct answers, because differences in the number of correct answers could be due to varying difficulty of the item sets taken by each respondent, for example. Instead, IRT scaling is used. IRT models differentiate between observable variables, such as a correct response to an item, and non-observable variables, such as the latent proficiency of a respondent. IRT scaling creates a common continuous scale that links the latent proficiency with the item difficulty via a probabilistic function. In PIAAC, a latent regression item response model for scaling, which encompassed calibration, population modeling, and the generation of plausible values, was used.

Item Calibration in PIAAC

In PIAAC, responses to literacy and numeracy items were dichotomously scored, and responses to PS-TRE items were either dichotomously or polytomously scored (with two or three correct answers).[14] In the calibration step, item parameters (item difficulty and item discrimination) were estimated using the two-parameter logistic model (2PL) for dichotomously scored responses, and the generalized partial credit model (GPCM) for polytomously scored responses. Responses to cognitive items using weighted data from 22 countries[15] were utilized to obtain international item parameters separately for each domain. Item calibration also took the different administration modes into account; a number of items common to both the paper and the computer mode shared the same parameter (Yamamoto et al., 2013b).

It was not necessary to drop any items from the PIAAC analysis. For a small number of assessment items, the international item parameters were not equal for all countries; this occurs when there is differential item functioning (DIF), i.e., the item functions differently in a specific country, for example, the item is harder or easier than in the other countries. Country-specific item parameters were estimated for cases with substantial deviations.[16] Per domain, country-specific parameters were estimated for only between 3% and 8% of the items (Yamamoto et al., 2013c, p. 19). According to Table 17.6 of the international technical report (Yamamoto et al., 2013c, p. 19), very few (only five of 166, i.e., 3 %) misfitting items were detected for Germany: three literacy and two numeracy items.[17]

14 One literacy item was scored polytomously.
15 Data from France and Russia were not included because their data delivery was delayed.
16 If items exhibited a comparable DIF in different countries, common item parameters for this group of countries were generated.
17 Appendix 17.1 lists a larger number of DIFs than in Table 17.6 (Yamamoto et al., 2013c). However, Germany remains one of the countries with the smallest number of DIFs.

In addition, a linked scale for trend measurement was produced, to provide comparable results for PIAAC with respect to IALS and ALL. This was accomplished in a combined analysis using data from all three surveys, in which item parameters were obtained from a concurrent item calibration (Yamamoto et al., 2013c).

Population Modeling and Plausible Values in PIAAC

Following the calibration of the item parameters with IRT, the resulting item parameters were used in a latent regression model in the population modeling process to obtain regression estimates. The plausible values methodology was used to improve the accuracy of the proficiency estimates for subpopulations and the overall population (Yamamoto et al., 2013c).

In the population model, latent proficiency was considered as the dependent variable. Predictor variables were both cognitive item responses as well as (basically all) background variables collected in the background questionnaire, which first underwent a principal component analysis. Both the principal component analysis and population modeling were carried out individually for each country, to account for national specifics (e.g., variables that explain more or less variance in the one or the other country).

To overcome the measurement error associated with the PIAAC matrix assessment design, and to obtain unbiased subpopulation and population estimates, an empirically derived distribution of proficiency values conditional on observed item response patterns and background variables was constructed for each respondent. Subsequently, plausible values were drawn at random from this posteriori distribution for each respondent. These plausible values are not test scores, but imputed values (which carry some random error variance) that have similarities with test scores and approximate an individual's proficiency; however, they are not suitable for computing a person's individual performance.

In PIAAC, ten plausible values per domain were independently drawn from the posteriori distribution for each respondent.[18] However, no plausible values for PS-TRE were imputed for respondents who were administered the paper assessment. Although each set of plausible values would be suitable to compute any statistic of interest, all sets of plausible values are needed to properly estimate the standard error of the respective statistic (Yamamoto et al., 2013c). This can be done with standard statistical software.

The Proficiency Scales and Levels

As described above, international calibration yielded a common, comparable scale for each assessment domain. These proficiency scales represent a continuum of competencies, which range from 0–500 points in PIAAC;[19] items located at the lower end of the scale are easier than items located at the higher end (Tamassia & Lennon, 2013). Both assessment items and respondents can be located on these scales by using a defined response probability (RP) value. In PIAAC, the RP value was set to 67%.[20] Thus, a respondent with a particular value on the PIAAC proficiency scale, for example 200, will successfully complete an item located at that same point (with an item difficulty of 200) with a probability of 67%. This respondent will solve a more difficult item (with a higher value on the scale) with a lower probability, and an easier item (with a lower value on the scale) with a higher probability.

In order to facilitate the interpretation of the proficiency scores, each scale was subdivided into various levels, defined by specific score-point ranges. The level boundaries, as given in Annex A.4, are identical for the domains literacy and numeracy, and somewhat different for PS-TRE, reflecting both the reduced number of items in PS-TRE as well as their greater difficulty, compared to the literacy and numeracy items (Yamamoto et al., 2013b). The international domain expert groups provided descriptions for each proficiency

18 There are no plausible values for literacy-related nonrespondents for any domain.
19 In theory, the scale ranges from $-\infty$ to $+\infty$.
20 Other surveys may use different RP values.

level, based on the underlying conceptual framework and the item characteristics at that level. These content-based level descriptions can be found in several OECD documents (OECD, 2013a, 2013b; Tamassia & Lennon, 2013) and, in German, in Zabal et al. (2013).

4.3.2 Data Confidentiality

The PIAAC survey explicitly considered ethical and privacy concerns from the beginning of the project. Survey ethics were an important subject in the interviewer manual and training. Data privacy concerns were also addressed.

During the course of data processing, at least elementary data confidentiality edits, such as eliminating any names, addresses, or other information allowing a direct identification of the respondents from their national databases, were automatically implemented in the Data Management Expert for all PIAAC countries. Beyond that, some countries implemented additional measures of data confidentiality to comply with their own, more restrictive data protection legislation. This was the case for Germany. The German National Center developed a data confidentiality strategy for the German PIAAC data that would offer German respondents appropriate protection while, at the same time, releasing as much data and detail as possible to data users.[21] This strategy was discussed and approved by a committee consisting of members of the German National Center, other GESIS experts, and the GESIS data protection officer. It laid out different data sets and their recipients, and specified different measures to be taken for these data sets.

The most complete and detailed data were needed by the international Consortium to carry out weighting and scaling. As mentioned in Section 4.1, a small set of confidentiality edits and variable suppressions (e.g., all open entry responses) was carried out prior to the submission of the German database to the international Consortium, and all recipient institutions were required to sign affidavits before being given access to these data.

The OECD data dissemination plan for PIAAC included releasing data via a web-based tool, the *Data Explorer*, and as international flat files (Public Use Files). For these data sets, Germany implemented a number of additional confidentiality edits and variable suppressions:[22] For example, only minimal regional information was provided (i.e., at a highly aggregated level, if at all). Age information was released only as a categorical and not as a continuous variable. Occupation and industry codes were made available at the two-digit level only. A number of other variables, such as information referring to countries, languages, income, household size, etc., were coarsened or top-coded.

At this stage, data confidentiality edits were defined by each country, but implemented in the national data sets by the international Consortium. While a requested suppression of variables by a country was always implemented, coarsening and top-coding of variables was only possible to a limited extent. A process was established, during which countries were invited to make proposals for additional coarsened or top-coded variables. From this pool of proposals, the OECD and the international Consortium selected and computed a set of variables that was considered to maximize data availability and comparability across countries; these were included in the international dissemination databases.[23] The German National Center made a number of proposals for coarsening and top-coding. Variables for which the proposed coarsenings were not implemented (and no appropriate alternative was offered) were suppressed in the German Public Use File.

21 In Germany, all target persons were informed about data confidentiality in a comprehensive data confidentiality statement that was sent to them together with the advance letter.

22 The German databases for the Data Explorer and the Public Use File are very similar, but we were less restrictive with the Data Explorer database because it does not provide direct access to the underlying database.

23 Variable names (e.g., in the Public Use File) ending with "_C" indicate a coarsened or top-coded variable.

4.3.3 Data Products and Tools

The OECD data dissemination plan included the public release of the PIAAC data, to make it accessible to a variety of data users (e.g., national governments, researchers, data analysts, or journalists). For this purpose, a microdata Public Use File was produced for each participating country. The Public Use Files were created from countries' preliminary final national files with implemented data confidentiality edits. All data files encompass the same original and derived *international* variables. Variables that were suppressed by a country contain missing values; *nationally* adapted variables or extensions are excluded. The Public Use Files for all countries (except Australia[24] and Cyprus) can be accessed without restrictions at the OECD website; they are available in SAS or SPSS format.[25]

In addition, in Germany, a Scientific Use File with a more comprehensive data set was produced. This Scientific Use File is stored at GESIS and is accessible to academic researchers only (study number: ZA5845, doi:10.4232/1.11865). Interested persons have to register with the GESIS Data Catalog and sign a data distribution contract specifically obliging them to adhere to data security laws.[26]

The German Scientific Use File includes detailed information for almost all background questionnaire variables, with only a few exceptions. For example, variables with information on the respondents' origin, such as country of birth, citizenship, and country in which the highest qualification was obtained, are coarsened. All German adaptations and extensions to the background questionnaire are included in this data set. Furthermore, some sampling and regional information is also released in this data set. The assessment data are supplemented by information on interviewer scorings of respondents' answers to paper-based core items. The User Guide (Perry & Helmschrott, 2014) provides more information on the German Scientific Use File (also accessible under the same link as the Scientific Use File).

The OECD has made several tools available that facilitate the analysis of the PIAAC data. These tools take the replicate weights and the plausible values used in PIAAC into account. The following tools are available at the OECD website:[27]

- The *Data Explorer*: This is an easy-to-use, web-based application that allows users to perform basic analyses (e.g., computation of means and percentages, with standard errors) and produce outputs (e.g., tables, graphics) from the PIAAC data. Users do not have access to the individual data records.
- The *International Database (IDB) Analyzer*: This is a software tool that produces SPSS code and allows basic analyses, such as the computation of means, percentages, percentiles, and linear regressions; standard errors can be calculated. This tool must be used with the Public Use Files in SPSS format. It is a product of IEA.[28]
- *SAS* and *STATA macros*: Macros for SAS and STATA are available for download at the OECD website. Both tools allow, for example, the computation of means, percentages, percentiles, as well as linear and logistic regressions.

24 Australian data can only be accessed directly from the Australian Bureau of Statistics (ABS).
25 http://www.oecd.org/site/piaac/publicdataandanalysis.htm
26 https://dbk.gesis.org/dbksearch/sdesc2.asp?no=5845
27 http://www.oecd.org/site/piaac/publicdataandanalysis.htm
28 It is also available at http://www.iea.nl/data.html.

4.4 Overall Quality

As this report has made clear, substantial efforts were undertaken in PIAAC, both at international and national levels, to achieve comparable data of high quality. A comprehensive set of standards and guidelines, based on best practice in survey and assessment methodology, was established for the PIAAC survey, accompanied by a thorough and continuous international quality assurance process that monitored all areas of survey implementation in each country and ensured that participating countries were provided with guidance and advice. The German National Center established the best possible implementation of the PIAAC standards and guidelines within the national constraints, and also carried out rigorous national monitoring of implementation processes and components. Although adhering to international standards was a central objective, considerable effort was invested in optimizing these nationally, additionally integrating national quality measures, as feasible.

In order to arrive at a final global evaluation of the quality of the national PIAAC data, the OECD and the Board of Participating Countries established a process of *data adjudication*. The objective of this adjudication was to determine whether the data were fit for public dissemination and use, and, if applicable, which limitations should apply to a release (for example, whether a cautionary footnote needed to be included). For the data adjudication, initial national quality reports were produced by the international Consortium. The Technical Advisory Group reviewed these and subsequently produced a report with recommendations for each participating country. The Board of Participating Countries made the final decision about inclusion and limitations to inclusion of the national databases (if any).

The following summary of the content and results of the data adjudication in PIAAC is based on Appendix 7 of the international technical report (OECD, 2013c, Appendix 7), but focuses on the evaluation of the German data.

Data adjudication focused on four key areas: (1) sampling, (2) coverage and nonresponse bias, (3) data collection, and (4) instrumentation.

For each of these areas, a set of quality indicators that reflected the key PIAAC requirements were selected and countries' compliance with these were evaluated. One of three possible outcomes was assigned: (a) *Pass*, relevant requirements completely met, (b) *Caution*, relevant requirements met to a reasonable extent, and (c) *Fail*, relevant requirements generally not met. A subsequent evaluation at the domain level focused on whether the quality of the domain was sufficient overall.

The evaluation of the sampling domain was based on the sampling plan, sample selection, sample weighting, and the sampling error. Although the evaluation of the German sampling met the requirements for most sampling indicators, a caution was assigned to this domain, because the selection probabilities were based on simulations and not exact calculations, due to the error in sample selection described in Section 3.1; the German sample remained, however, a probability-based sample. As a consequence of this error, the German design effect due to sampling variance was also somewhat larger than intended (German design effect was 1.22), but it remained at an adequate level. In all, four PIAAC countries received a caution for sampling.

The evaluation of the coverage and nonresponse bias was based on the results of the population coverage, the design-weighted response rate, and the basic and extended NRBAs. Germany met the requirements for all NRBAs, as well as the requirements for the coverage indicator, with a total noncoverage rate of 2.5%. For Germany, as for nearly all other participating countries, the international Consortium concluded that the weighting adjustments were effective in reducing the potential for bias. Overall, 20 out of the 25 countries, including Germany, received a caution for this area (generally with the addendum "bias low", as for Germany; in two cases, with the addendum "bias minimal") because their

response rates were below 70%, regardless of the outcomes of the NRBAs.[29] This caution was given to reflect a potential for minimal or low levels of bias in the outcome statistics for response rates below this benchmark (OECD, 2013c, Appendix 7).

Two indicators were evaluated for the domain data collection: (a) field validation/rechecks, and (b) staffing, training, management/monitoring. Germany and the majority of other PIAAC countries met the (full or reduced) requirements for this quality domain. Only four countries were assigned a caution.

The requirements regarding the quality of the assessment data and the background questionnaire data, translation, coding, and scoring, and item non-response were evaluated for the quality domain instrumentation. Germany fulfilled all requirements in this domain, as did all but one of the participating countries (one country was assigned a fail).

The adjudication process went beyond the compliance with standards in these four key areas, and focused especially on the overall data quality, with a view to the intended data use. Thus, the final evaluation concerns the *fitness-for-use* of the PIAAC data, in terms of whether the data have the quality required for the intended applications.

All data were accepted without limitation in the international data base and international reporting (OECD, 2013a), with the exception of one country[30] whose results are still of a preliminary nature and for which a cautionary note is included.

From a national point of view, PIAAC was an unusually complex survey and challenging to implement. However, the joint efforts and the unusual commitment of all those involved—from the interviewers, programmers, coders, to survey organization home-office staff, GESIS and national experts, and numerous others—have made PIAAC in Germany a very special and, in our view, successful project.

29 Please note: Twenty-five countries are listed here because response rates were calculated and reported separately for England and Northern Ireland.

30 Russian Federation

Annex

A.1 PIAAC Field Test

All countries participating in PIAAC had to conduct a field test between April and June 2010.[1] The main goal of the PIAAC field test was to test the survey instruments, procedures, and certain aspects of the design, in preparation for the main survey. This included (a) evaluating the psychometric properties of items and scales, as well as the comparability and timing of the background questionnaire and assessment instruments; (b) specifically for the assessment instruments, evaluating the equivalence of computer and paper modes as well as the linking to the previous surveys, IALS and ALL; (c) testing the newly developed computer-based delivery platform; and (d) checking the appropriateness of sampling and survey operations. The field test was regarded as a general dress rehearsal for the main survey, both at an international as well as at the national level. The final instrumentation for the main survey was selected by the international Consortium, based on analyses of field test data.

This annex gives an overview of the central features of the field test. We will focus on the major differences between the field test and main survey (more detailed information on the field test is provided in the international technical report; OECD, 2013c). Furthermore, we will summarize the German implementation of sampling and survey operations for the field test.

Instrumentation

The international Consortium developed the survey instruments for the field test hand-in-hand with the assessment and questionnaire frameworks produced under the guidance of international domain expert groups.[2] The survey and assessment design for the field test was specified by the international Consortium to address key field test objectives.

The cognitive assessment in the field test and the main survey were both self-administered by the respondents. However, the assessment design for the field test was completely different from that of the main survey, because the new computer-based implementation of the assessment was to be critically examined in the field test. Specifically, the field test assessment design tested the equivalence of the paper versus the computer delivery, as well as the equivalence of the IALS and ALL linking items (IALS/ALL versus PIAAC). To test for potential mode effects, the field test assessment design ensured that a sufficiently large number of computer-literate respondents (i.e., respondents who had previously worked with a computer) carried out the assessment on paper. This was achieved by *randomly* assigning a certain proportion of respondents to either the paper branch or the computer branch. The field test analyses did not identify any significant mode effects (OECD, 2013b).

The computer-based assessment in the field test was not adaptive, as in the main survey, but used a balanced block order with a fixed item order within each block (Yamamoto et al., 2013a). Eight assessment booklets were implemented in the paper-based assessment: two literacy and two numeracy booklets (per domain, the booklets included the same items but

1 Of the 26 countries that participated in the field test, all except Chile and Portugal also completed the main survey.

2 Countries also contributed to this process. More information on this is provided in Chapter 2.

in a different order; in addition, core items were included), and four reading components booklets (each with a separate set of exercises). Respondents were administered a total of two booklets: one of the four literacy or numeracy booklets and one of the four reading components booklets.

The field test also explored how best to assess the computer familiarity required for the computer-based assessment. Prior to the assessment, respondents were asked a few questions regarding their ICT skills, to determine whether they would be able to take the computer-based assessment. Respondents who did not have sufficient computer experience could choose to receive a short tutorial on how to use a computer mouse. Both of these elements were later excluded from the main survey design.

For the background questionnaire, the field test implemented a block design (contrary to the main survey). The advantage of this was that more questions could be tested without placing too much burden on each respondent. All respondents were administered a core block of questions as well as one of four subsets of questions randomly. The core block referred to general information, education, current employment status, employment history, and background information. The random blocks were either about (a) skills used at home, (b) reading and writing skills used at work and at home, (c) numeracy and ICT skills used at work and at home, or (d) additional information such as attitudes and personal activities.

Another focus of the field test was to examine whether the Job Requirement Approach[3] questions could also yield valid and reliable information for unemployed respondents. Therefore, the field test assessed how well unemployed respondents could recall retrospective information about their skill use at their most recent job within the last 12 months. As a result of field test analyses, it was determined that these questions could be reliably administered to respondents who had been unemployed within the last year; this was accordingly implemented in the main survey (Allen et al., 2013). As in the main survey, the field test background questionnaire was conducted in CAPI mode.

As described in Section 2.4, all instruments and survey materials were translated by countries for the field test. The field test results therefore also served to check the quality of the translations as well as the functioning and comparability of all national instruments.

Following completion of the field test, the instruments were analyzed by the international Consortium. Final questions for the background questionnaire of the main survey were selected on the basis of analytical priorities, content, data quality, and timing considerations. For the assessment, item selection focused on the psychometric and linking properties of the items, comparability, and construct coverage.

PIAAC Computer Delivery Platform

An additional function of the field test was to test the implementation and stability of the new computer delivery platform described in Section 2.5. At a national level, this included testing the integration of the virtual machines into the national case management software. Various problems occurred both with the international technical components and their delivery and, for some countries (including Germany), with the integration into the national case management system. Although many of these bugs were identified and corrected prior to the data collection, some remaining technical problems had to be patched during the field period. The technical problems observed in the field test were, in most instances, corrected prior to the main survey, so that only minor technical issues were encountered during the main survey.

Sampling and Survey Operations

The field test also gave countries the opportunity to test and evaluate their sampling and survey operations. The majority of the international standards and guidelines for these areas were valid for both the main survey and the field test. However, the field test allowed for

3 See Section 2.2

more degrees of freedom, provided that these did not have a negative impact on the overall field test goals.

According to the international PIAAC standards for the field test, a representative sample was not required, but it was recommended to test probability-based sampling. A minimum of 1 500 completed cases had to be obtained during a three-month data collection period, in order to have a sufficiently large sample for analytical purposes.

Given these facts, we implemented a probability-based sampling design in Germany that was supplemented by a quota design during the last quarter of the data collection period, due to an insufficient number of completed cases from the random sample. As in the main survey, for the random sample, a registry-based stratified and clustered sampling design was implemented (see Section 3.1 for more details). In contrast to the main survey, the field test design had three selection stages: (1) selection of five federal states, (2) selection of municipalities, and (3) selection of individuals. The field test sampling design was restricted to five federal states,[4] chosen so that they would represent all of Germany on a smaller scale (e.g., suitably represent small, medium, and large municipalities).

Overall, 68 municipalities were selected with a total of 80 sample points, and 40 addresses were selected for each sample point. The gross sample consisted of 3 198 cases.[5] A reserve sample of another twelve addresses per sample point was available for the replacement of ineligibles.[6] Less than a third of the reserve sample was released; the final gross sample thus consisted of 3 455 cases.

Overall, 1 185 net interviews were obtained from the registry sample. These cases had to be supplemented by another 400 interviews from a quota sample. The quota sample was selected on the basis of a three-dimensional matrix of age, gender, and education. In total, 469 quota interviews were completed. From this pool of quota interviews, 400 interviews were subsequently selected, such that the final distributions of age, gender, and education best matched the German Microcensus statistics.

The field test offered the opportunity to establish and test best practice fieldwork procedures and to develop study materials and interviewer documentation in preparation for the main survey. In Germany, 65 interviewers were hired for the field test, and they all underwent a five-day interviewer training (as was the case for the main survey interviewers). Interviewer feedback from the field test debriefing sessions proved valuable for improving processes for the main survey.

Considering the very challenging standards regarding response rates in PIAAC (a minimum of 50% if analyses showed low nonresponse bias, 70% as the general goal), Germany implemented an incentive experiment in the field test in order to identify the most promising incentive strategy for the main study. This experiment tested three conditional incentives. One of the following conditions was randomly assigned to the sampled persons with a ratio of 40:40:20: (a) 50 euros in cash, (b) 25 euros in cash, and (c) a 10-euro commemorative coin, engraved with a motif of the 2006 Soccer World Cup. Results of this experiment indicated that the 50 euro conditional incentive performed better than the other two conditions. Therefore, this incentive was used in the main survey.

In addition, two separate versions of the brochure for target persons were tested; one included a reference to PISA, the other did not contain such a reference.[7] The purpose was to test whether including a reference to PISA had an effect on survey participation. Results showed no difference between the two versions.

---

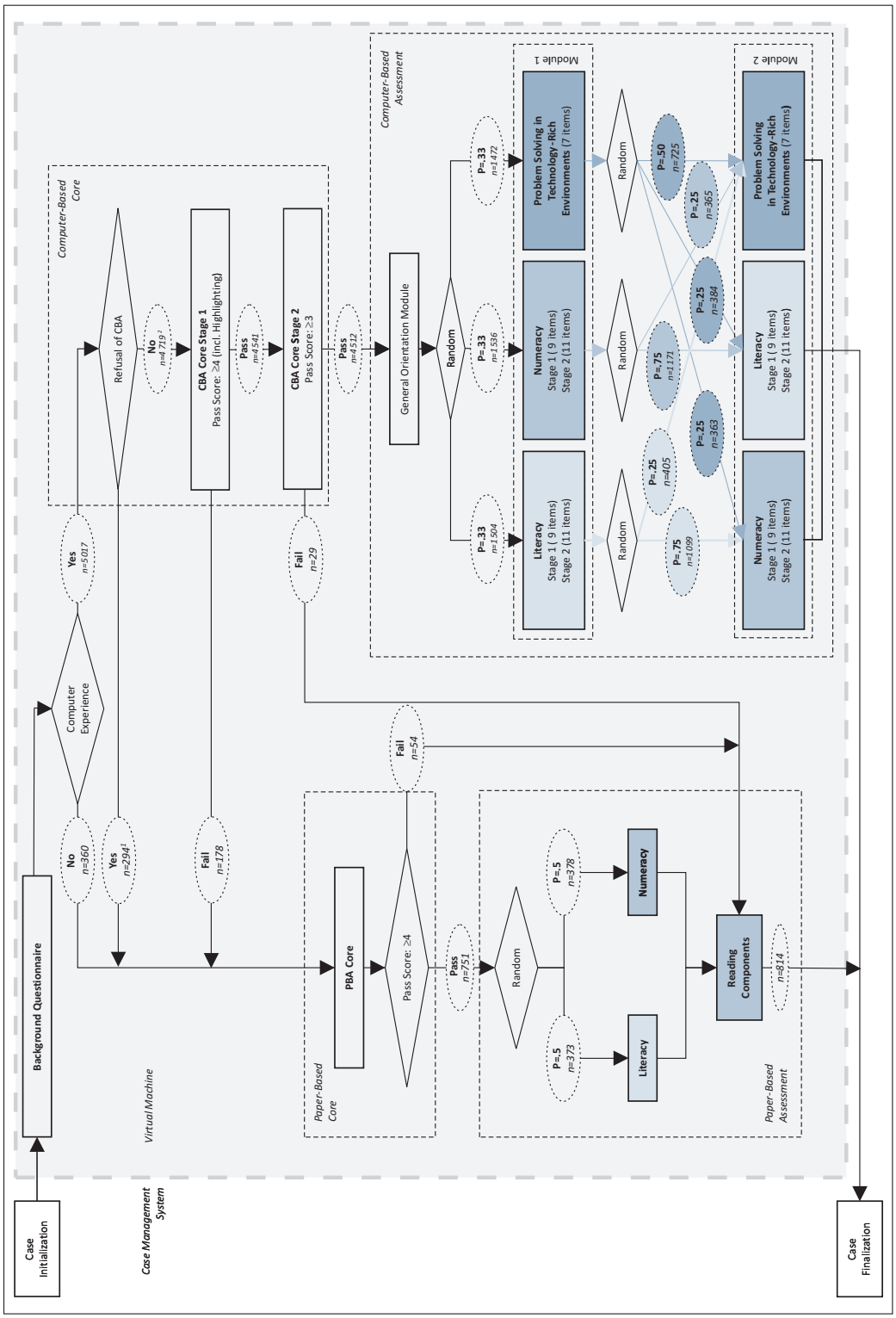
4 Bavaria, Hamburg, Saxony, Schleswig-Holstein, and Thuringia

5 One municipality did not deliver the minimum number of cases.

6 The approach of replacing cases identified as ineligibles with new target persons from a reserve sample was allowed in the field test, contrary to the main survey.

7 The brochures were pre-allocated to sampled persons at a ratio 50:50 and sent to respondents along with the advance letter.

A.2.1 Detailed PIAAC Interview Workflow



Notes. Cp. Martin et al. (2013, p. 174) and OECD (2011b). Breakoffs are excluded at the stage at which they occurred and the subsequent stages. There were twelve breakoffs in the computer-based assessment (five of them were technical breakoffs). Reported number of cases after the reading components module excludes respondents with missing reading components booklets. CBA = computer-based assessment. P = allocation probability. *n* = number of cases.

[1] One breakoff to the CBA branch, which was routed to the PBA branch, is included, because it was treated equivalently to refusals of the CBA branch. [2] One technical breakoff is excluded.

Figure A.2.1. Detailed PIAAC interview workflow

A.2.2 Examples of German Reading Components Items

Teil 1. Anleitung: Kreisen Sie das Wort ein, das zu dem Bild passt.

3.

Ohr Kamm Lippe Tor

Note. German example item translated from the international English OECD version (2013b, p. 25).

Figure A.2.2a. Reading components example item for word meaning

Anleitung: Kreisen Sie JA ein, wenn der Satz einen Sinn ergibt. Kreisen Sie NEIN ein, wenn der Satz keinen Sinn ergibt.

4. Drei Mädchen aßen den Bus.	**JA**	**NEIN**
5. Der Mann fuhr das grüne Auto.	**JA**	**NEIN**
7. Ein bequemes Kissen ist weich und Stadt.	**JA**	**NEIN**

Note. German example item translated from the international English OECD version (2013b, p. 25).

Figure A.2.2b. Reading components example item for sentence processing

A.2.3 List of Linking Units

Table A.2.3a. Literacy Linking Units

		Literacy	
Unit N°	Unit Name	Assessment mode	Linking to
300	Employment Ad	Paper and Computer	ALL
301	SGIH	Paper and Computer	ALL
302	Election Results	Paper and Computer	ALL, IALS
304	Contact Employer	Computer only	ALL, IALS
305	TMN AntiTheft	Paper and Computer	ALL
306	CANCO	Paper and Computer	ALL, IALS
307	MEDCO	Computer only	ALL, IALS
308	Baltic Stock Market	Computer only	ALL
309	Generic Medicines	Paper and Computer	ALL
310	Memory Training	Paper and Computer	ALL
311	Dutch Women	Computer only	ALL, IALS
313	International Calls	Paper and Computer	ALL
315	Mexican Distances	Computer only	IALS

Table A.2.3b. Numeracy Linking Units

		Numeracy	
Unit N°	Unit Name	Assessment mode	Linking to
600	Election Results	Paper and Computer	ALL
601	Bottles	Paper and Computer	ALL
602	Price Tags	Paper and Computer	ALL
604	Gas Gauge	Paper and Computer	ALL
605	Photo	Computer only	ALL
606	Solution	Computer only	ALL
607	TV	Computer only	ALL
608	Tree	Computer only	ALL
610	Compound Interest	Paper only	ALL
611	Temp Scale	Computer only	ALL
612	Dioxin	Computer only	ALL
613	Logbook	Computer only	ALL
614	Watch	Paper and Computer	ALL
615	Candles	Paper and Computer	ALL
617	Map	Computer only	ALL
618	Sixpack	Paper and Computer	ALL
619	Tiles	Computer only	ALL
620	Inflation	Paper and Computer	ALL
622	Classified	Computer only	ALL
623	Wine	Paper and Computer	ALL
624	BMI	Paper and Computer	ALL

A.2.4 International PIAAC Scripts

The German technical implementation made use of all the international PIAAC scripts, which offered the following functionalities:

- Virtual machine remote control
  - Start the virtual machine
  - Shut down the virtual machine

- Case management and data access
  - Import case parameters
  - Start case
  - Resume case
  - Get case status
  - Export interview data

- Handle maintenance and administrational requests
  - Install patches
  - Data recovery

A.3.1 PIAAC Advance Letter

TNS Infratest Sozialforschung GmbH
Landsberger Straße 284
80687 München
Kostenfreie Hotline: 0800 - 100 1425
(Mo-Do 9.00 bis 17.00 Uhr;
Fr 9.00 bis 15.00 Uhr)
piaac@tns-infratest.com

Herr Gernemitmacher
Adresszeile1
Adresszeile2

München, August 2011

PIAAC – Was brauchen wir, um Alltag oder Beruf erfolgreich zu meistern?

Sehr geehrter Herr Gernemitmacher,

Sie wurden ausgewählt, an der PIAAC-Befragung teilzunehmen. PIAAC ist eine internationale Studie, die zeitgleich in 25 Ländern durchgeführt wird. Ziel der Studie ist es, herauszufinden, welche Fertigkeiten wir in unserem Leben bisher erworben haben und was wir brauchen, um den alltäglichen und beruflichen Anforderungen zu begegnen.

Für PIAAC werden in Deutschland von 54 Millionen Erwachsenen im Alter von 16 bis 65 Jahren etwa 5.000 Personen befragt. Die Auswahl dieser Personen erfolgt zufällig aus den Einwohnermeldeämtern. Selbstverständlich ist eine Teilnahme freiwillig. Jedoch sind die Ergebnisse nur dann aussagekräftig, wenn möglichst alle ausgewählten Personen teilnehmen. Daher sind auch *Sie* persönlich und *Ihre* Teilnahme besonders wichtig. Als Zeichen unserer Anerkennung erhalten Sie nach Durchführung der Befragung **50,- €** in bar.

In Deutschland wird PIAAC von TNS Infratest Sozialforschung im Auftrag von GESIS – Leibniz-Institut für Sozialwissenschaften durchgeführt. GESIS ist eine wissenschaftliche Forschungseinrichtung, die vom Bund und den Ländern finanziert wird.

In den nächsten Wochen wird sich Frau Marianne Muster mit Ihnen persönlich in Verbindung setzen. Sie ist die für Sie vorgesehene Interviewerin von TNS Infratest und zeigt Ihnen gerne ihren Interviewerausweis.

Selbstverständlich halten wir bei PIAAC den Datenschutz ein; genauere Informationen dazu entnehmen Sie bitte der beiliegenden Erklärung zum Datenschutz. Weitere Informationen zu PIAAC finden Sie auch im Faltblatt.

Wir freuen uns sehr, wenn Sie an dieser wichtigen Befragung teilnehmen und bedanken uns schon im Voraus ganz herzlich für Ihre Unterstützung.

Mit freundlichen Grüßen

Günter Steinacker
Projektleiter TNS Infratest Sozialforschung
TNS Infratest Sozialforschung GmbH
Landsberger Straße 284
80687 München
E-Mail: piaac@tns-infratest.com

Prof. Dr. Beatrice Rammstedt
Projektleiterin von PIAAC Deutschland
GESIS – Leibniz-Institut für Sozialwissenschaften
Postfach 12 21 55
68072 Mannheim
E-Mail: info-piaac@gesis.org

A.3.2 PIAAC Flyer

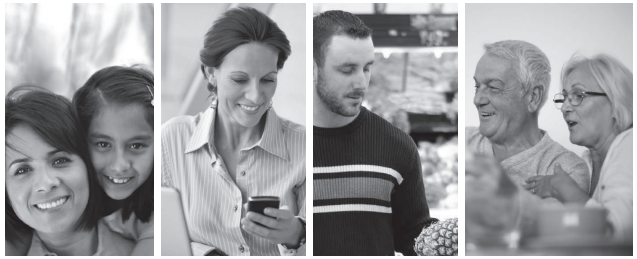
gesis
Leibniz-Institut für Sozialwissenschaften

PIAAC – Was brauchen wir, um Alltag oder Beruf erfolgreich zu meistern?

Eine internationale Studie zur Untersuchung von Alltagsfertigkeiten Erwachsener

tns infratest

Fotos: iStockphoto



PIAAC
Programme for the International Assessment of Adult Competencies

gesis
Leibniz-Institut für Sozialwissenschaften

Im Überblick: PIAAC ...

- steht für „Programme for the International Assessment of Adult Competencies"
- untersucht die Alltagsfertigkeiten Erwachsener im internationalen Vergleich
- liefert eine wissenschaftliche Grundlage für politische Entscheidungen und gesellschaftliche Veränderungen
- wird 2011/12 in 25 Ländern durchgeführt
- befragt in jedem Land etwa 5.000 Personen zwischen 16 und 65 Jahren
- wird im Auftrag der OECD (Organisation für wirtschaftliche Entwicklung und Zusammenarbeit) durchgeführt
- wird in Deutschland
  - vom Bundesministerium für Bildung und Forschung (BMBF) unter Beteiligung des Bundesministeriums für Arbeit und Soziales (BMAS) finanziert
  - von GESIS – Leibniz-Institut für Sozialwissenschaften geleitet
  - von den Interviewern von TNS Infratest durchgeführt
- **wird mit IHNEN ein Erfolg!**

OECD

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

GEFÖRDERT VOM
Bundesministerium für Arbeit und Soziales

PIAAC
Programme for the International Assessment of Adult Competencies

Haben Sie noch Fragen?

Auf der Internetseite von PIAAC (www.gesis.org/piaac) finden Sie ausführliche Informationen rund um die Studie.

Für Sie als Teilnehmerin bzw. Teilnehmer hat das Umfrageinstitut TNS Infratest eine kostenfreie Hotline eingerichtet. Der Projektleiter, Herr Günter Steinacker, und sein Team stehen Ihnen für Fragen gerne zur Verfügung.

Kostenfreie Telefon-Hotline für Befragte: 0800 – 100 14 25

(Mo-Do 9.00 bis 17.00 Uhr; Fr 9.00 bis 15.00 Uhr)

TNS Infratest Sozialforschung GmbH
Landsberger Straße 284
80687 München
www.tns-infratest-sofo.com
Mail: piaac@tns-infratest.com

Herausgeber

GESIS – Leibniz-Institut
für Sozialwissenschaften
Postfach 12 21 55
68072 Mannheim
www.gesis.org/piaac
Mail: info-piaac@gesis.org

PIAAC
Programme for the International Assessment of Adult Competencies

gesis
Leibniz-Institut für Sozialwissenschaften

PIAAC – Eine internationale Studie zur Untersuchung von Alltagsfertigkeiten Erwachsener

Was ist PIAAC eigentlich?

PIAAC steht für „Programme for the International Assessment of Adult Competencies" und ist eine internationale Studie zur Untersuchung von Alltagsfertigkeiten Erwachsener. Sie wird im Auftrag der OECD zeitgleich in 25 Ländern durchgeführt.

Was sind die Ziele von PIAAC?

In Deutschland liegen bislang kaum Informationen über die Fertigkeiten von Erwachsenen vor. Genau hier soll PIAAC ansetzen und z.B. folgende Fragen beantworten:

- Welche Alltagsfertigkeiten sind in unserer Gesellschaft vorhanden und wo stehen wir im Vergleich zu anderen Ländern?
- Wie werden diese Kenntnisse im Alltag und im Beruf genutzt und wie können sie gefördert werden?
- Wie geübt sind wir im Umgang mit neuen Technologien?

Warum sind die Ergebnisse von PIAAC wichtig?

PIAAC liefert Informationen, die für wegweisende Entscheidungen in der Bildungs- und Arbeitsmarktpolitik verwendet werden sollen. Die in PIAAC erhobenen Daten bilden damit eine Grundlage für die aktive Gestaltung unserer Zukunft.

Wann gibt es erste Ergebnisse?

Erste Ergebnisse werden Ende 2013 der Öffentlichkeit vorgestellt.

Worüber werden Sie befragt?

Neben allgemeinen Fragen zu Ihrer Person, wie z.B. Alter, Familienstand oder beruflicher Werdegang, erhalten Sie einige Aufgaben aus verschiedenen Bereichen des Alltags: So könnten Sie gebeten werden, auf einem Stadtplan den kürzesten Weg zwischen zwei Orten zu bestimmen oder auf dem Beipackzettel eines Medikaments abzulesen, wie lange dieses eingenommen werden darf.

Wie sieht es mit dem Datenschutz aus?

Selbstverständlich werden bei PIAAC alle Auflagen des Datenschutzes eingehalten. Alle personenbezogenen Daten werden streng vertraulich behandelt.

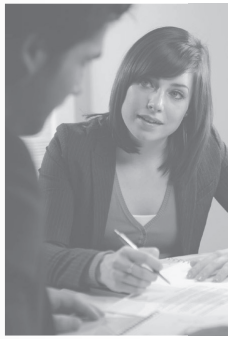
Wann findet die Befragung statt?

PIAAC wird von Sommer 2011 bis Frühjahr 2012 durchgeführt. GESIS – Leibniz-Institut für Sozialwissenschaften hat TNS Infratest mit der Durchführung der Befragung beauftragt. Eine Interviewerin bzw. ein Interviewer von TNS Infratest wird sich persönlich mit Ihnen in Verbindung setzen und einen für Sie passenden Termin vereinbaren. Das Interview dauert im Durchschnitt 90 Minuten.

Warum wurden Sie ausgewählt?

Sie wurden zufällig aus 54 Millionen Menschen in der Altersgruppe der 16- bis 65-Jährigen in Deutschland ausgewählt. Nur durch eine rein zufällige Auswahl von Personen sind verlässliche Rückschlüsse auf unsere gesamte Bevölkerung möglich.

Warum sollten Sie an PIAAC teilnehmen?



Damit die Ergebnisse für Deutschland aussagekräftig sind, ist es entscheidend, dass *alle* ausgewählten Personen teilnehmen – deshalb ist auch **Ihre** Teilnahme besonders wichtig.

Es ist egal, ob Sie berufstätig sind oder nicht, ob Sie sich noch in der Ausbildung befinden oder schon in Rente sind: Uns geht es um Ihre persönliche Erfahrung, denn die Vielfalt in unserer Gesellschaft soll sich auch bei PIAAC widerspiegeln.

Selbstverständlich ist die Teilnahme an PIAAC freiwillig. Sie können aber maßgeblich zum Erfolg der Studie in Deutschland beitragen.

Bitte machen auch **Sie** bei PIAAC mit!

Unser Dankeschön an Sie:

Wir wissen, wie wertvoll Ihre Zeit ist. Sie erhalten von uns daher 50,- Euro als Dankeschön für Ihren Zeitaufwand.

PIAAC
Programme for the International Assessment of Adult Competencies

A.3.3 PIAAC Brochure

gesis
Leibniz-Institut für Sozialwissenschaften

PIAAC – Was brauchen wir, um Alltag oder Beruf erfolgreich zu meistern?

tns infratest

Eine internationale Studie zur Untersuchung von Alltagsfertigkeiten Erwachsener

PIAAC
Programme for the International Assessment of Adult Competencies

gesis
Leibniz-Institut für Sozialwissenschaften

Gliederung

PIAAC

Programme for the International Assessment of Adult Competencies

Sehr geehrte Damen und Herren,

Sie wurden ausgewählt an PIAAC teilzunehmen. PIAAC ist eine international vergleichende Studie zu Fertigkeiten von Erwachsenen und wird zeitgleich in 25 Ländern weltweit durchgeführt.

Mit dieser Broschüre möchten wir Sie über die Studie und den Ablauf der Befragung informieren.

Was Sie als Teilnehmerin bzw. Teilnehmer von PIAAC aufbringen sollen, ist lediglich Ihre Zeit. Genauer gesagt handelt es sich um etwa 90 Minuten, in denen Sie gemeinsam mit dem Interviewer unsere Fragen beantworten.

Ziel von PIAAC ist es, herauszufinden, welche Fertigkeiten wir im Laufe unseres Lebens erworben haben. Ob beim Gang zum Supermarkt, zum Arzt, zur Bank oder im Job – überall treffen wir auf kleinere oder größere Aufgaben, die es zu meistern gilt. Inwieweit nutzen wir unsere Fertigkeiten, um solchen alltäglichen und beruflichen Anforderungen erfolgreich zu begegnen? Wie können Bildungsangebote verbessert und Kenntnisse optimal gefördert werden, damit unsere Gesellschaft auch in Zukunft erfolgreich sein kann? Helfen Sie mit, dass PIAAC hierauf die richtigen Antworten geben kann!

Wir, die nationale Projektleiterin und der Präsident der GESIS – Leibniz-Institut für Sozialwissenschaften, bitten Sie, an unserer Studie teilzunehmen und so zum Erfolg von PIAAC beizutragen.

Für Ihre Unterstützung bedanken wir uns schon heute ganz herzlich.

Prof. Dr. Beatrice Rammstedt
Projektleiterin von PIAAC Deutschland

Prof. Dr. York Sure
Präsident der GESIS

GESIS – Leibniz-Institut für Sozialwissenschaften

gesis
Leibniz-Institut für Sozialwissenschaften

4

Allgemeine Informationen zu PIAAC

Was ist PIAAC eigentlich?

PIAAC ist eine internationale Studie zur Untersuchung von Alltagsfertigkeiten Erwachsener und steht für „Programme for the International Assessment of Adult Competencies". Zu Alltagsfertigkeiten gehören das Lesen und Verstehen von Texten, der Umgang mit Zahlen oder mit dem Internet. Es geht um Aufgaben, die uns häufig in Alltag begegnen, wie z.B. das Überschlagen eines Sonderangebotes oder die Informationssuche in Zeitungen.

Warum gibt es PIAAC und was sind die Ziele von PIAAC?

PIAAC ermöglicht erstmalig Aussagen darüber, welche Fertigkeiten Erwachsene in Deutschland haben und wie sie diese im privaten und beruflichen Alltag nutzen. Die Ergebnisse liefern wichtige Informationen für Entscheidungen in der Bildungs- und Arbeitsmarktpolitik. Nur wenn Schule, Ausbildung und Weiterbildung auf die Anforderungen der modernen Wirtschaft und der neuen Technologien vorbereiten, können nachfolgende Generationen erfolgreich im Beruf und im Leben bestehen.

„Es ist wichtig, dass Kinder etwas lernen, womit sie später einmal etwas anfangen können."

PIAAC Programme for the International Assessment of Adult Competencies

5

In welchen Ländern wird PIAAC durchgeführt?

Weltweit nehmen 25 Länder in Europa, Asien, Amerika sowie Australien an der Studie teil.

Was bringt ein internationaler Vergleich der Ergebnisse?

PIAAC gibt einen Überblick darüber, wie die Bevölkerung der teilnehmenden Länder mit Anforderungen des Alltags und Berufs umgeht. Durch den Vergleich der Ergebnisse können die einzelnen Länder voneinander lernen und so beispielsweise Möglichkeiten finden, Aus- und Weiterbildungssysteme weiterzuentwickeln.

gesis
Leibniz-Institut für Sozialwissenschaften

6

Allgemeine Informationen zu PIAAC

Wer steht hinter PIAAC und wer finanziert die Studie?

Auf internationaler Ebene wird die Studie von der Organisation für wirtschaftliche Entwicklung und Zusammenarbeit (OECD) organisiert. In Deutschland wird PIAAC vom Bundesministerium für Bildung und Forschung (BMBF) unter Beteiligung des Bundesministeriums für Arbeit und Soziales (BMAS) finanziert.

Für die Durchführung in Deutschland ist GESIS – Leibniz-Institut für Sozialwissenschaften verantwortlich. Mit der Befragung der Teilnehmerinnen und Teilnehmer wurde das Umfrageinstitut TNS Infratest von GESIS betraut.

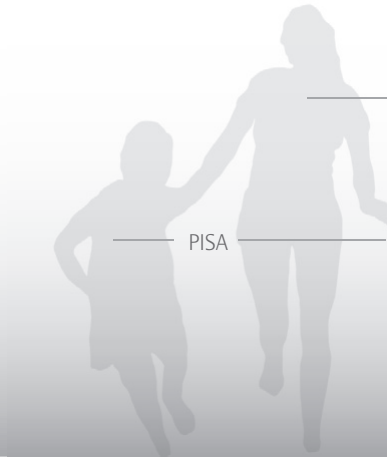
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT OECD

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

GEFÖRDERT VOM Bundesministerium für Arbeit und Soziales

tns infratest

„Technik vereinfacht unsere Arbeitsabläufe – dafür bildet man sich gerne weiter."



PISA

PIAAC Programme for the International Assessment of Adult Competencies

7

Hat PIAAC etwas mit PISA zu tun?

Ja, denn ähnlich wie PISA hat auch PIAAC das Ziel, Fertigkeiten international vergleichend zu untersuchen. Darüber hinaus wird PIAAC, ebenso wie PISA, von der OECD international organisiert. Im Gegensatz zur PISA-Studie, bei der hauptsächlich Schülerinnen und Schüler befragt werden, liegt der Schwerpunkt bei PIAAC auf der Befragung Erwachsener. Außerdem werden bei PISA Schulleistungen untersucht, während bei PIAAC Alltagsfertigkeiten beleuchtet werden. PIAAC und PISA sind zwei eigenständige Studien, die unabhängig voneinander durchgeführt werden.

Wann gibt es erste Ergebnisse?

Erste Ergebnisse der Studie für Deutschland und die weiteren Teilnehmerländer sollen Ende 2013 veröffentlicht werden.

— PIAAC

„Nicht jedes Sonderangebot ist wirklich günstig. Man muss schon genau hinsehen."

gesis
Leibniz-Institut für Sozialwissenschaften

8

„Es ist schön seine Kenntnisse im Alltag zu nutzen – vor allem wenn man dabei Spaß hat."

„Es ist praktisch Termine schnell bestätigen zu können."

„Ausbildung muss sich an den wirklichen Anforderungen des Lebens orientieren."

„Auf einer Baustelle muss man sich auf die Arbeit der anderen verlassen können - jeder muss dazu beitragen, damit alle erfolgreich sind."

„Heute sind im Beruf ganz andere Fähigkeiten gefragt als vor 20 Jahren."

PIAAC Programme for the International Assessment of Adult Competencies

Fragen zu Ihrer Teilnahme an PIAAC

9

Wer wird befragt und wie wurden Sie ausgewählt?

Bei PIAAC werden Erwachsene im Alter zwischen 16 und 65 Jahren befragt. Mithilfe von Daten aus den Einwohnermeldeämtern wurden Sie aus 54 Millionen Menschen in Deutschland zufällig ausgewählt. Natürlich ist Ihre Teilnahme freiwillig. Mit Ihrer Teilnahme tragen Sie aber maßgeblich zum Gelingen der Studie bei: Nur wenn möglichst alle der ausgewählten Personen an der Studie teilnehmen, sind die gewonnenen Ergebnisse aussagekräftig.

Wie wird der Datenschutz eingehalten?

Selbstverständlich ist die Einhaltung des Datenschutzes das oberste Gebot von PIAAC. Alle personenbezogenen Angaben werden streng vertraulich behandelt und nicht an Dritte weitergegeben. Aus den Daten oder Ergebnissen der Studie ist nicht ersichtlich, von welcher Person die Angaben gemacht wurden. Weiterführende Informationen zum Thema Datenschutz finden Sie auf unserer Internetseite unter www.gesis.org/piaac.

Wie lange dauert eine Befragung?

Die Befragung dauert durchschnittlich 90 Minuten. Im Zeitraum von August 2011 bis März 2012 wird sich eine Interviewerin oder ein Interviewer hierfür mit Ihnen in Verbindung setzen und einen für Sie passenden Termin für die Befragung vereinbaren.

gesis
Leibniz-Institut für Sozialwissenschaften

10

Fragen zu Ihrer Teilnahme an PIAAC

Worüber werden Sie befragt?

Zunächst werden Ihnen allgemeine Fragen zu Ihrer Person gestellt. Dazu gehören zum Beispiel Angaben zu Ihrer Ausbildung, Ihrem beruflichen Werdegang oder Fragen zu Aktivitäten in Ihrer Freizeit. Anschließend erhalten Sie Aufgaben, die Sie aus dem täglichen Leben kennen. Im untenstehenden Informationskasten finden Sie einige Beispiele für mögliche Aufgaben.

Beispiele für mögliche Aufgaben

- Sie sollen einen Beipackzettel lesen und beantworten, ob dieses Medikament für Kinder geeignet ist.
- Sie sollen den kürzesten Weg zu einem Restaurant auf dem Stadtplan einzeichnen.
- Sie sollen verschiedene Internetseiten vergleichen und beurteilen, welche Seite geeignet ist, um ein Rezept zu finden.
- Sie sollen anhand einer Preisliste überschlagen, wie viel der Einkauf im Supermarkt kosten wird.

PIAAC untersucht mit diesen Aufgaben Fertigkeiten aus verschiedenen Bereichen: dem Lesen und Verstehen von Texten sowie dem alltäglichen Umgang mit Zahlen. Wenn Sie über Computerkenntnisse verfügen, kann es auch um die allgemeine Verwendung des Computers gehen, z.B. darum E-Mails zu schreiben oder das Internet zu nutzen.

„Wir brauchen Azubis, die gut auf ihren Beruf vorbereitet sind."

PIAAC Programme for the International Assessment of Adult Competencies

11

Was bringt Ihnen die Teilnahme an PIAAC?

Mit Ihrer Teilnahme leisten Sie einen entscheidenden Beitrag zur positiven Entwicklung unseres Bildungssystems und unserer Arbeitsmarktchancen. Sie tragen somit zu der langfristigen Sicherung unserer wirtschaftlichen und gesellschaftlichen Wettbewerbsfähigkeit bei. Die Teilnahme eines jeden Einzelnen ist für die Studie wichtig, da nur so für Deutschland ein umfassender und aussagekräftiger Überblick der Alltagsfertigkeiten von Erwachsenen erstellt werden kann.

Als Entschädigung für Ihren Zeitaufwand erhalten Sie ein finanzielles Dankeschön in Höhe von **50,–** Euro.

Wo finden Sie weitere Informationen?

Auf der Internetseite von PIAAC finden Sie unter **www.gesis.org/piaac** ausführliche Informationen rund um die Studie. Außerdem hat das Umfrageinstitut TNS Infratest für Sie als Teilnehmerin bzw. Teilnehmer eine kostenfreie Hotline (**0 800 – 100 14 25**) eingerichtet. Der Projektleiter, Herr Günter Steinacker, und sein Team stehen Ihnen für Fragen gerne zur Verfügung.

„Es macht Freude, sich jeden Tag zu informieren und so auf dem Laufenden zu bleiben."

gesis
Leibniz-Institut für Sozialwissenschaften

PIAAC

Impressum

Ansprechpartner für die Befragung:

TNS Infratest Sozialforschung GmbH
Landsberger Straße 284
80687 München
www.tns-infratest-sofo.com
Mail: piaac@tns-infratest.com

Kostenfreie Telefon-Hotline für Befragte: *0 800 – 100 14 25*
(Mo-Do 9.00 bis 17.00 Uhr; Fr 9.00 bis 15.00 Uhr)

Herausgeber:

GESIS – Leibniz-Institut für Sozialwissenschaften
Postfach 12 21 55
68072 Mannheim
www.gesis.org/piaac
Mail: info-piaac@gesis.org

www.gesis.org

Fotos: iStockphoto LP

A.3.4 Key Facts About the PIAAC Supplementary Sample

- Commissioned by Prof. Dr. Heike Solga, WZB (Berlin Social Science Center), Berlin
- Target population: 26 to 55 year olds in former East Germany
- Sample selection according to the German PIAAC sample design (at stage 2) in 56 of 320 PIAAC sample points (located in East Berlin, Brandenburg, Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt, and Thuringia)
- Data collection by TNS Infratest in parallel to PIAAC Germany 2012
- Gross sample size: 1 174 cases
- Two main data collection phases: 12 October to 31 December 2011, 28 January to 27 February 2012
- Four re-issue phases: 7 December 2011 to 13 February 2012, 28 January to 27 February 2012, 24 February to 26 March 2012, 9 to 26 March 2012
- Interviewers: 30 of the 129 PIAAC interviewers
- Administration according to standards applied for the German PIAAC sample (e.g., same instruments, interviewers, documents for respondents, incentives, fieldwork monitoring, and control procedures, etc.); supervision by the German National Center
- Realized sample: 560 cases
- Fieldwork response rate:[8] 47%

- Final disposition codes:

Final disposition code	*n*	%
1 Completed interview	544	46.3
4 Refusal sample person[1]	406	34.6
5 Refusal other person[1]	15	1.3
6 Refusal due to time constraints[1]	24	2.0
7 Language problem	8	0.7
- Initial, with age/gender collected	7	
- In assessment	1	
9 Learning/mental disability (with age/gender collected)[1]	6	0.5
12 Hearing impairment[1]	1	0.1
14 Speech impairment[1]	1	0.1
15 Physical disability[1]	2	0.2
16 Other disability[1]	2	0.2
17 Other reasons (unspecified), such as sickness or unusual circumstances	14	1.2
- Initial	12	
- In assessment	2	
18 Death[1]	1	0.1
21 Non-contact with household[1]	70	6.0
22 Non-contact with sample person[1]	20	1.7
24 Sample person temporarily absent/unavailable during field period[1]	7	0.6
31 Sample person moved into institution[1]	3	0.3
32 Sample person moved outside country[1]	11	0.9
33 Sample person moved to other community (within Germany)[1]	7	0.6
34 Sample person moved, new residence unknown[1]	26	2.2
35 Invalid address[1]	6	0.5
Total	**1 174**	**100.0**

Note. n = number of cases.

[1] Only initial.

8 Response rate = completed background questionnaire/gross sample

- International data cleaning and processing, weighting and scaling by the international Consortium (independently of the PIAAC sample)
- Activities at the German National Center after data collection:
  - application of same data management activities as for the PIAAC sample (as described in Section 4),
  - combination of the realized oversample of 560 cases with 3 440 cases from the PIAAC sample (*only* 26 to 55 year olds); thus, the *combined supplementary data set* consists of 4 000 cases of 26 to 55 year olds,
  - specification of weighting variables and weighting benchmarks for the *combined supplementary sample* (weighting variables chosen for each weighting step were generally kept identical to those used for the PIAAC sample, see Section 3.4; Microcensus benchmark data were obtained for the target group of 26 to 55 year olds),
  - submission of the *combined supplementary sample* of 4 000 cases to the international Consortium for weighting and scaling,
  - final data checks upon receipt of final *combined supplementary data set*, including weights and plausible values from the international Consortium, and
  - delivery to the researcher team.
- Note: This additional sample is neither part of the international Public Use File nor the PIAAC Scientific Use File for Germany.

A.3.5 Checklist for PIAAC Interview

- Laptop and accessories (computer mouse, cable, and charged battery)
- PIAAC case folder
- Interviewer ID badge
- Copies of the advance letter
- Copies of the confidentiality statement
- Flyer
- Callback cards
- Brochures
- Interviewer manual
- Tape recorder and spare batteries
- 50 euros in cash as incentive for respondent
- 2 sets of paper booklets (for assessment) each containing:
  - Core Booklet (blue)
  - Task Booklet 1, literacy (yellow)
  - Task Booklet 2, numeracy (cream)
  - Exercise Booklet A, reading components (green)
- Show card booklet(s)
- Ruler
- Calculator
- 2 photos
- Note pads
- Pens (black)
- Interviewer booklet

A.3.6 PIAAC Case Folder

TNS Infratest	Aktenmappe	Die Aktenmappe ist vertraulich, sie darf Dritten in keiner Weise zugänglich gemacht werden.
PIAAC 01	101 1	1234567
Projekt Welle	Listennr. lfd. Nr.	Internationale ID

Bitte befragen Sie:

Frau /Herr
Vorname Nachname
Straße Hausnummer
PLZ Ort Ortsteil

Altersgruppe
20 bis 39 Jahre

Bitte Fragen 1 bis 5 für **alle** Zielpersonen beim ersten Aufsuchen der Adresse, **vor dem ersten** persönlichen Kontaktversuch, ausfüllen.

1.) Verfügt das Haus über eine Gegensprechanlage?

- ☐ Ja
- ☐ Nein

2.) In welcher Art von Gebäude wohnt die Zielperson? Bitte nur eine Kategorie auswählen!

- ☐ Landwirtschaftliches Wohngebäude
- ☐ Freistehendes Ein-/Zweifamilienhaus
- ☐ Ein-/Zweifamilienhaus als Reihenhaus oder Doppelhaus
- ☐ Wohnhaus mit 3 bis 4 Wohnungen
- ☐ Wohnhaus mit 5 bis 8 Wohnungen
- ☐ Wohnhaus mit 9 oder mehr Wohnungen (höchstens 8 Stockwerke, kein Hochhaus)
- ☐ Hochhaus (9 oder mehr Stockwerke)
- ☐ Sonstiges Haus / Gebäude, und zwar: _____

3.) Wie beurteilen Sie den Zustand des Hauses?

- ☐ in sehr gutem Zustand
- ☐ in gutem Zustand
- ☐ etwas renovierungsbedürftig
- ☐ stark renovierungsbedürftig

Bei den beiden folgenden Fragen geht es einfach um **Ihre beste Einschätzung**. Bitte auch diese Fragen vor dem ersten persönlichen Kontaktversuch ausfüllen und später nicht mehr korrigieren.

4.) Zu welcher Schicht gehört Ihrer Meinung nach die Zielperson? Vermutlich zur...

- ☐ Unterschicht
- ☐ Arbeiterschicht
- ☐ Mittelschicht
- ☐ Oberen Mittelschicht
- ☐ Oberschicht

5.) Über welchen Schulabschluss verfügt Ihrer Meinung nach die Zielperson? Vermutlich...

- ☐ Niedriger Schulabschluss
- ☐ Mittlerer Schulabschluss
- ☐ Höherer Schulabschluss

TNS Infratest	Aktenmappe	Die Aktenmappe ist vertraulich, sie darf Dritten in keiner Weise zugänglich gemacht werden.

PIAAC 01	101	1	1234567
Projekt	Welle	Listennr.	lfd. Nr.	Internationale ID

Bitte befragen Sie:

Altersgruppe
20 bis 39 Jahre

Frau /Herr
Vorname Nachname
Straße Hausnummer
PLZ Ort Ortsteil

Bitte JEDEN Kontakt und JEDEN Kontaktversuch EINZELN notieren !

Nr	Wochen-tag	Datum	Uhrzeit	Kontakt-art	Kontakt-ergebnis	(Neuer) Termin / Notizen zum Kontakt(versuch), Notizen zur Adresse und für Nachbearbeitung auf Seite 3 (unten) eintragen!
1						Bitte immer zuerst Seite 1 ausfüllen!
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

Kontaktart:
1. persönlich
2. telefonisch d. mich
3. telefonisch d. ZP
4. über Gegensprechanlage
5. Info durch Infratest
6. Email/Andere

Finales Kontaktergebnis
Ergebnis mit kleinstem Buchstaben

für 3 bis 6, 17, 22, 24:

für 7 bis 16:

Unbedingt: Erläuterung zum finalen Kontaktergebnis

Alter: Geschlecht: M W

Ich bestätige die Richtigkeit der auf dieser Aktenmappe gemachten Angaben:

Abrechnungsnr. Name Datum Unterschrift

Seite 2

Liste der Kontaktergebnisse

1	A	Vollständiges Interview
3	F	Abbruch
4	F	Zielperson nicht zum Interview bereit, weil:
5	F	Andere Person im Haushalt lässt Interview nicht zu, weil:
6	F	Zielperson aus Zeitgründen nicht zum Interview bereit
7	B	Sprachproblem
8	B	Lese- und Schreibschwierigkeiten
9	B	Lernbehinderung / Geistige Behinderung
12	C	Hörbehinderung
13	C	Blindheit/Sehbehinderung
14	D	Sprachbehinderung
15	C	Körperliche Behinderung
16	C	Sonstige Behinderung
17	E	Sonstige Gründe (keine näheren Angaben), z.B. Krankheit oder außergewöhnliche Umstände
18	C	Zielperson (ZP) verstorben
21	H	Im HH niemanden angetroffen
22	G	ZP nicht angetroffen, aber im HH jemanden angetroffen
24	H	ZP vorübergehend abwesend/während der Feldzeit nicht erreichbar
31	C	ZP lebt in Anstalt und nicht in Privathaushalt
32	C	ZP umgezogen ins Ausland
33	D	ZP umgezogen außerhalb der Gemeinde (innerhalb Deutschlands)
34	D	ZP umgezogen, neuer Wohnort unbekannt
35	D	Adresse falsch, existiert nicht (mehr)

Zwischenergebnis, nicht final:

41	Terminvereinbarung
42	ZP angetroffen, aber noch kein finales Kontaktergebnis
43	ZP umgezogen innerhalb der Gemeinde: Adresse ermitteln und nachverfolgen! (=> Notizfeld **Adresse** nutzen!)
44	Pausiertes Interview
90	Technisches Problem

Notizen zur **Adresse** / zur Befragung:

_____

_____

_____

_____

_____

Wichtige Informationen für die evtl. **Nachbearbeitung** der Adresse:
WICHTIG: Informationen in COMPASS Fragebogen übertragen!

_____

_____

_____

_____

Telefonnummer(n):

_____

Wir wünschen Ihnen gutes Gelingen bei der Bearbeitung dieses anspruchsvollen Projekts.

Ihre TNS Infratest Feldorganisation
Die Projektleitung bei TNS Infratest Sozialforschung
Das PIAAC Projektteam bei GESIS

Als Notfalltelefon steht Ihnen die Projektleitung von TNS Infratest Sozialforschung unter der folgenden aus dem Festnetz kostenfreien Telefonnummer zur Verfügung:

0800 / 1001425

TNS Infratest	Aktenmappe	Die Aktenmappe ist vertraulich, sie darf Dritten in keiner Weise zugänglich gemacht werden.

Zu den Interviews:

Wurde das Set an papierbasierten Heften genutzt?

☐ Ja, Seriennummer:

☐ Nein

Haben Sie das Interview mit dem Aufnahmegerät mitgeschnitten?

☐ Ja ☐ Nein, abgelehnt von der Zielperson

☐ Nein, nicht gefragt ☐ Nein, nicht mehr nötig (2 Aufnahmen liegen vor)

Wurde für den Hintergrundfragebogen ein Dolmetscher eingesetzt?

☐ Nein ☐ Ja

Wenn ja, bitte Beziehung zur Zielperson angeben:

Über die Ausfälle *mit Kontakt zur Zielperson / zum Haushalt:*

Welche Sprache spricht die Zielperson heute zu Hause am häufigsten?

☐ Deutsch ☐ Weiß nicht

☐ Türkisch ☐ Sonstige, und zwar:

☐ Russisch

Checkliste zum Abschließen einer Adresse*:*

		erledigt
1.)	Aktenmappe wurde vollständig ausgefüllt	☐
2.)	Falls papierbasierte Hefte zum Einsatz gekommen sind: Hefte in Aktenmappe abgelegt für Rücksendung (bitte stets alle vier Hefte eines Sets zurückschicken)	☐
3.)	Adresse in COMPASS abgeschlossen	☐

A.3.7 Results of Interviewer Information About Interview Situation

Interview situation	% *n* = 5 375
Presence of additional person	20.0
Additional person assisted in background questionnaire	
- Yes	4.2
- No	15.7
Additional person assisted in assessment	
- Yes	3.1
- No	16.9
Respondent understood questions	
- Never	0.4
- Almost never	0.9
- Now and then	7.9
- Often	23.6
- Very Often	67.1
Respondent asked for clarification in assessment	18.6
Events that occurred during the interview	
- Respondent held conversation with someone else	5.0
- Respondent answered a phone call, text message or email	7.6
- Respondent was looking after children	2.2
- Respondent was undertaking domestic tasks	0.2
- TV/radio/game console/music player was in use in the respondent‘s vicinity	2.8
- Respondent was interrupted by some other activity, task or event	3.9
Respondent complained that assessment was taking too long	8.6
Room of assessment	
- Living/dining room	65.8
- Kitchen	20.9
- Bedroom	0.3
- Entrance	0.3
- Hallway or corridor	0.6
- Office	4.2
- Other space in the household	3.6
- Other space outside of the household	4.4

Notes. The number of cases (n = 5 375) includes completed cases for which an interviewer answered the questions about the interview setting. No answers were obtained for literacy-related nonrespondents (LRNR) as well as for a few cases with technical problems.

A.4 Boundaries of Proficiency Levels

Table A.4a. Boundaries of Proficiency Levels for Literacy and Numeracy

Level	Range of score points
Below level 1	< 176
1	176–225
2	226–275
3	276–325
4	326–375
5	≥ 376

Table A.4b. Boundaries of Proficiency Levels for PS-TRE

Level	Range of score points
Below level 1	< 241
1	241–290
2	291–340
3	≥ 341

References

Allen, J., Van der Velden, R., Helmschrott, S., Martin, S., Massing, N., Rammstedt, B., . . . Von Davier, M. (2013). The Development of the PIAAC Background Questionnaire. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 3, pp. 1-48). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Andersson, R., & Olsson, A.-K. (1999). *Fields of Education and Training. Manual*. Luxembourg: Eurostat.

Antoni, M., Drasch, K., Kleinert, C., Matthes, B., Ruland, M., & Trahms, A. (2010). *Arbeiten und Lernen im Wandel: Teil I: Überblick über die Studie. März 2011 (2. Aktualisierte Fassung des Berichtes vom Mai 2010)*. FDZ Methodenreport 05/2010. Nürnberg: FDZ.

Aust, F., Gilberg, R., Hess, D., Kersting, A., Kleudgen, M., & Steinwede, A. (2012). *Methodenbericht NEPS Startkohorte 6: Haupterhebung 2010/2011*. Bonn: Institut für angewandte Sozialwissenschaft GmbH.

Aust, F., Gilberg, R., Hess, D., Kleudgen, M., & Steinwede, A. (2011). *Methodenbericht NEPS Etappe 8: Befragung von Erwachsenen. Haupterhebung 1. Welle 2009/2010*. Bonn: Institut für angewandte Sozialwissenschaft GmbH.

Bethlehem, J. G. (2002). Weighting Nonresponse Adjustments Based on Auxiliary Information. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 275-288). New York: Wiley.

Biemer, P. P. (2010). Overview of Design Issues: Total Survey Error. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (2 ed., pp. 27-57). Bingley, United Kingdom: Emerald Group Publishing.

Blohm, M., Harkness, J., Klein, S., & Scholz, E. (2003). *Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2002.* ZUMA-Methodenbericht 2003/12. Mannheim: ZUMA.

Blohm, M., & Koch, A. (2013). Respondent Incentives in a National Face-to-Face Survey: Effects on Outcome Rates, Sample Composition and Fieldwork Efforts. *Methoden, Daten, Analysen. Zeitschrift für empirische Sozialforschung, 7*(1), 89-122. doi: 10.12758/mda.2013.004

Carstens, R., & Daniel, T. (2013). Data Management Procedures. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 13, pp. 1-33). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Carstens, R., Daniel, T., & Gonzalez, E. (2013). International Database and Data Analysis Tools. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 23, pp. 1-16). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Cox, L. H. (1987). A Constructive Procedure for Unbiased Controlled Rounding. *Journal of the American Statistical Association, 82*(398), 520-524. doi: 10.1080/01621459.1987.10478456

European Social Survey. (2012). *ESS5-2010 Documentation Report. The ESS Data Archive*. (3.0 ed.). Bergen: Norwegian Social Science Data Services. Retrieved March 2014, from http://www.europeansocialsurvey.org/docs/round5/survey/ESS5_data_documentation_report_e03_0.pdf

European Social Survey. (2013). *ESS6-2012 Documentation Report. The ESS Data Archive* (1.3 ed.). Bergen: Norwegian Social Science Data Services. Retrieved March 2014, from http://www.europeansocialsurvey.org/docs/round6/survey/ESS6_data_documentation_report_e01_3.pdf

Felstead, A., Gallie, D., Green, F., & Zhou, Y. (2007). *Skills at work, 1986 to 2006*. Retrieved from http://kar.kent.ac.uk/4845/1/2006_Skills_Survey-FinalDraft3.pdf

Ferrari, A., Wayrynen, L., Behr, D., & Zabal, A. (2013). Translation, Adaptation, and Verification of Test and Survey Materials. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 4, pp. 1-28). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Gabler, S., & Ganninger, M. (2010). Gewichtung. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 143-164). Wiesbaden: VS Verlag für Sozialwissenschaften.

Gal, I., Alatorre, S., Close, S., Evans, J., Johansen, L., Maguire, T., . . . Tout, D. (2009). *PIAAC Numeracy: A Conceptual Framework.* OECD Education Working Paper No. 35. Paris: OECD Publishing.

Hamburgisches Weltwirtschaftsinstitut. (2010). Stocks of Irregular Migrants: Estimates for Germany. Retrieved September 06, 2011, from http://irregular-migration.net//typo3_upload/groups/31/3.Database_on_IrregMig/3.2.Stock_Tables/Germany_Estimates_IrregularMigration_Dec10.pdf

Harter, R., Eckman, S., English, N., & O'Muircheartaigh, C. (2010). Applied Sampling for Large-Scale Multistage Area Probability Designs. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (2 ed., pp. 169-197). Bingley, United Kingdom: Emerald Group Publishing.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). *Applied Survey Data Analysis*. Boca Raton: CRC Press.

International Labour Office. (2012). *International Standard Classification of Occupations ISCO-08: Structure, Group Definitions and Correspondence Tables.* Geneva: International Labour Organization. Retrieved from http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_172572.pdf

Jadoul, R., Plichart, P., Bogaerts, J., Henry, C., & Latour, T. (2013). The TAO Platform. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 9, pp. 1-26). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Jones, S., Gabrielsen, E., Hagston, J., Linnakylä, P., Megherbi, H., Sabatini, J., . . . Vidal-Abarca, E. (2009). *PIAAC Literacy: A Conceptual Framework.* OECD Education Working Paper No. 34. Paris: OECD Publishing.

Kirsch, I., & Yamamoto, K. (2013). Assessment Design. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 1, pp. 1-17). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Kleinert, C., Matthes, B., & Jacob, M. (2008). *Die Befragung „Arbeiten und Lernen im Wandel": Theoretischer Hintergrund und Konzeption.* IAB Forschungsbericht 5/2008. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit.

Koch, A. (1997). ADM-Design und Einwohnermeldeamtsregister-Stichprobe: Stichprobenverfahren bei mündlichen Bevölkerungsumfragen. In S. Gabler & J. H. P. Hoffmeyer-Zlotnik (Eds.), *Stichproben in der Umfragepraxis* (pp. 99-116). Opladen: Westdeutscher Verlag.

Lennon, M. L., & Tamassia, C. (2013). The Development of the PIAAC Cognitive Instruments. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 2, pp. 1-28). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Library of Congress Network Development and MARC Standards Office. (2013). Codes for the Representation of Names of Languages. Codes arranged alphabetically by alpha-3/ISO 639-2 Code. Retrieved March 18, 2014, from http://www.loc.gov/standards/iso639-2/php/code_list.php

Martin, S., Zabal, A., Helmschrott, S., Ackermann, D., Massing, N., Rammstedt, B., & Häder, S. (2013). Qualitätssicherung, Design und Datenqualität. In B. Rammstedt (Ed.), *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich – Ergebnisse von PIAAC 2012* (pp. 167-182). Münster: Waxmann.

Massing, N., Ackermann, D., Martin, S., Zabal, A., & Rammstedt, B. (2013). Controlling Interviewers' Work in PIAAC – the Programme for the International Assessment of Adult Competencies. In P. Winker, N. Menold & R. Porst (Eds.), *Interviewers' Deviations in Surveys. Impact, Reasons, Detection and Prevention* (pp. 117-130). Frankfurt am Main: Peter Lang.

Microm MARKET & GEO. (2011). *Datenhandbuch*. Unpublished document.

Mohadjer, L., Krenzke, T., & Van de Kerckhove, W. (2013a). Indicators of the Quality of the Sample Data. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 16, pp. 1-30). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Mohadjer, L., Krenzke, T., & Van de Kerckhove, W. (2013b). Sampling Design. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 14, pp. 1-31). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Mohadjer, L., Krenzke, T., Van de Kerckhove, W., & Hsu, V. (2013c). Survey Weighting and Variance Estimation. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 15, pp. 1-28). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Montalvan, P., & Lemay, M. (2013a). Field Operations. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 10, pp. 1-38). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Montalvan, P., & Lemay, M. (2013b). Quality Control Monitoring Activities. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 11, pp. 1-12). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Notter, P., Arnold, C., Von Erlach, E., & Hertig, P. (2006). *Lesen und Rechnen im Alltag: Grundkompetenzen von Erwachsenen in der Schweiz.* Neuchâtel: Bundesamt für Statistik.

OECD. (1999). *Classifying Educational Programmes. Manual for ISCED-97. Implementation in OECD Countries*. Retrieved from http://www.oecd.org/edu/1841854.pdf

OECD. (2010a). *PIAAC Reducing Nonresponse Bias and Preliminary Nonresponse Bias Analysis. PIAAC(2010_03)NonResponseBias_15March10.docx*. Unpublished document.

OECD. (2010b). *PIAAC Technical Standards and Guidelines*. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

OECD. (2011a). *PIAAC Conceptual Framework of the Background Questionnaire Main Survey*. Retrieved from http://www.oecd.org/site/piaac/PIAAC%282011_11%29MS_BQ_ConceptualFramework_1%20Dec%202011.pdf

OECD. (2011b). *PIAAC Main Study Virtual Machine – Part I. PIAAC(2011-02)MasterVM_Part_I.doc.* Unpublished document.

OECD. (2011c). *PIAAC Weighting and Variance Estimation Plan. PIAAC_NPM (2011_6) Weighting and Variance Estimation Plan.doc*. Unpublished document.

OECD. (2013a). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. Paris: OECD Publishing.

OECD. (2013b). *The Survey of Adult Skills – Reader's Companion*. Paris: OECD Publishing.

OECD. (2013c). *Technical Report of the Survey of Adult Skills (PIAAC)*. Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

OECD. (2013d). Territorial Grids of OECD Member Countries. Retrieved March 18, 2014 from http://www.oecd.org/governance/regional-policy/42392313.pdf. Website: http://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?Dataset=REG_LAB_TL3

OECD, & Statistics Canada. (2000). *Literacy in the Information Age. Final Report of the International Adult Literacy Survey*. Paris: OECD Publishing.

Perry, A., & Helmschrott, S. (2014). *User Guide for the German PIAAC Scientific Use File*. Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften.

Rammstedt, B. (2013). *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich – Ergebnisse von PIAAC 2012*. Münster: Waxmann.

Rouet, J.-F., Bétrancourt, M., Britt, M. A., Bromme, R., Graesser, A. C., Kulikowich, J. M., . . . Van Oostendorp, H. (2009). *PIAAC Problem Solving in Technology-Rich Environments: A Conceptual Framework.* OECD Education Working Paper No. 36. Paris: OECD Publishing.

Sabatini, J. P., & Bruce, K. M. (2009). *PIAAC Reading Components: A Conceptual Framework.* OECD Education Working Paper No. 33. Paris: OECD Publishing.

Statistics Canada, & OECD. (2005). *Learning a Living. First Results of the Adult Literacy and Life Skills Survey*. Paris: OECD Publishing.

Statistisches Bundesamt. (2004). *Demographische Standards. Methoden – Verfahren – Entwicklungen. Eine gemeinsame Empfehlung des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute e.V. (ADM), der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI) und des Statistischen Bundesamtes*. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2008). *Bevölkerung und Erwerbstätigkeit: Ausländische Bevölkerung – Ergebnisse des Ausländerzentralregisters*. Wiesbaden: Statistisches Bundesamt.

Stoop, I. A. L., Billiet, J., Koch, A., & Fitzgerald, R. (2010). *Improving Survey Response: Lessons Learned From the European Social Survey*. Chichester: John Wiley & Sons.

Tamassia, C., & Lennon, M. L. (2013). PIAAC Proficiency Scales. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 21, pp. 1-18). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Tamassia, C., Lennon, M. L., & Yamamoto, K. (2013). Scoring Reliability Studies. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 12, pp. 1-10). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

UNESCO Institute for Statistics. (2006). *International Standard Classification of Education ISCED 1997*. UNESCO Institute for Statistics (UNESCO-UIS). Retrieved from http://www.uis.unesco.org/Library/Documents/isced97-en.pdf

United Nations Statistics Division. (2013a). Detailed Structure and Explanatory Notes ISIC Rev. 4. Retrieved July 12, 2013, from http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27

United Nations Statistics Division. (2013b). Standard Country or Area Codes for Statistical Use. Retrieved March 18, 2014, from http://unstats.un.org/unsd/methods/m49/m49alpha.htm

Upsing, B., Goldhammer, F., Schnitzler, M., Baumann, R., Johannes, R., Barkow, I., . . . Wagner, M. (2013a). Development of the Cognitive Items. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 5, pp. 1-9). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Upsing, B., Goldhammer, F., Schnitzler, M., Baumann, R., Johannes, R., Barkow, I., . . . Wagner, M. (2013b). Development of the Integrated Computer Platform. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 8, pp. 1-14). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Wasmer, M., Scholz, E., & Blohm, M. (2010). *Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2008.* GESIS-Technical Reports. Bonn: GESIS – Leibniz-Institut für Sozialwissenschaften.

Wasmer, M., Scholz, E., Blohm, M., Walter, J., & Jutz, R. (2012). *Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2010.* GESIS-Technical Reports. Köln: GESIS – Leibniz-Institut für Sozialwissenschaften.

Yamamoto, K., Khorramdel, L., & Von Davier, M. (2013a). Proficiency Scale Construction. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 19, pp. 1-22). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Yamamoto, K., Khorramdel, L., & Von Davier, M. (2013b). Scaling Outcomes. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 18, pp. 1-28). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Yamamoto, K., Khorramdel, L., & Von Davier, M. (2013c). Scaling PIAAC Cognitive Data. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC)* (Chapter 17, pp. 1-33). Paris: OECD. Retrieved from http://www.oecd.org/site/piaac/surveyofadultskills.htm

Zabal, A., Martin, S., Klaukien, A., Rammstedt, B., Baumert, J., & Klieme, E. (2013). Grundlegende Kompetenzen der erwachsenen Bevölkerung in Deutschland im internationalen Vergleich. In B. Rammstedt (Ed.), *Grundlegende Kompetenzen Erwachsener im internationalen Vergleich – Ergebnisse von PIAAC 2012* (pp. 31-76). Münster: Waxmann.

List of Persons who Contributed to PIAAC Germany

The following persons were involved in the implementation of PIAAC in Germany. We thank everyone for the enormous effort and care that went into this ambitious work. We are also indebted to all members of the international Consortium for their patience and excellent support.

GESIS – Leibniz Institute for the Social Sciences

German National Center

Scientific Team:
Prof. Dr. Beatrice Rammstedt (NPM)
Anouk Zabal (Assistant NPM)
Daniela Ackermann
Susanne Helmschrott
Dr. Débora Maehler
Silke Martin (Data Manager)
Natascha Massing
Dr. Anja Perry

Secretariate and assistance:
Susanne Boetsch
Julia Khorshed
Maria Kreppe-Aygün
Catharina Zimmermann

Student assistants:
Anna Bachsleitner
Jule Bauckhage
Christian Behrendt
Thomas Brinker
Eva-Maria David
Apollonia Goll
Malte Reents
Christina Röhrig
Anna Schädler
Teresa Sluk

Additional valuable support from:
GESIS-Mannheim IT Department
PR Department
GESIS Administration

GESIS Experts

Dr. Dorothée Behr (Translation)
Michael Blohm (Survey Operations, Nonresponse)
PD Dr. Siegfried Gabler (Sampling)
Dr. Matthias Ganninger (Sampling)
Alfons Geis (Occupation Coding)
Dr. Sabine Häder (Sampling Manager)
Prof. Dr. Jürgen Hoffmeyer-Zlotnik (Education)
Dr. Katharina Kinder-Kurlanda (Data Archiving)
Achim Koch (Survey Operations)
Dr. Jan-Philipp Kolb (Sampling and Statistics)
Dr. Timo Lenzner (Linguistics)
Rolf Porst (Pretesting)
Peter Prüfer (Pretesting)
Dr. Silke Schneider (Education)
Martina Wasmer (Background Questionnaire)
Oliver Watteler (Data Archiving)
Dr. Heike Wirth (Data Privacy)

TNS Infratest Sozialforschung

Survey Organisation

Günter Steinacker (Project Director)
Sarah Schmidt (Assistant Project Director)
Ingo Leven (Project Director Field Test)

Numerous PIAAC interviewers, fieldwork supervisors, and other home office staff

DIPF (German Institute for International Educational Research)

National IT Coordination

National IT Team:
Ingo Barkow (National IT-Coordinator)
Robert Baumann (Assistant National IT-Coordinator)
Roland Johannes
Dr. Heiko Rölke

Administrative Project Assistance:
Sabrina Hermann
Carolin Ziegler

IEA DPC

Scoring / Occupation and Industry Scoring

Guido Martin (Lead Coder)
Stephanie Vocke (Lead Scorer)

Numerous DPC scorers and coders

Translators

Brita Dorer (Background Questionnaire)
Dr. Paul Kußmaul (Background Questionnaire)

Beate Staib (PS-TRE, Interviewer Manual)
Hella Beister (Literacy, Numeracy)

National Experts

Domain Experts

Prof. Dr. Annelies Blom (Nonresponse)
Prof. Dr. Aiso Heinze (Numeracy)
Dr. Corinna Kleinert (Background Questionnaire)
Prof. Dr. Detlev Leutner (PS-TRE)
Prof. Dr. Jürgen Schupp (Background Questionnaire)
Prof. Dr. Petra Stanat (Literacy)

Additional Experts

Prof. Dr. Cordula Artelt
Prof. Dr. Frank Goldhammer
Dr. Jan Hochweber
Prof. Dr. Frauke Kreuter
Dr. Johannes Naumann
Prof. Dr. Manfred Prenzel
Dr. Jean-Paul Reeff
Dr. Simon Wiederhold

National Scientific Advisory Board

Prof. Dr. Heike Solga (Chair)
Prof. Dr. Jürgen Baumert
Prof. Dr. Eckhard Klieme
Prof. Dr. Ursula Staudinger
Prof. Dr. Christof Wolf
Prof. Dr. Ludger Wößmann

Representatives from the Federal Ministry of Education and Research:
Dr. Thomas Greiner
Dr. Doerte Treuheit

Representatives from the Federal Ministry of Labor and Social Affairs:
Ulrike Fröhlich
Andreas Henkes

Permanent guests:

Prof. Dr. York Sure-Vetter
(President of GESIS – Leibniz Institute for the Social Sciences)

Udo Michallik
(Secretary General of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany)

List of Figures

List of Tables

List of Abbreviations

ABS	Australian Bureau of Statistics
AES	Adult Education Survey
ALL	Adult Literacy and Life Skills Survey
ALLBUS	Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (German General Social Survey)
ALWA	Arbeiten und Lernen im Wandel (Working and Learning in a Changing World)
BQ	Background questionnaire
BRR	Balanced repeated replication
CAL	Final weight
CAPI	Computer-assisted personal interview
CBA	Computer-based assessment
CRP	Centre de Recherche Public Henri Tudor (Public Research Centre Henri Tudor)
DIF	Differential item functioning
DIPF	Deutsche Institut für Internationale Pädagogische Forschung (German Institute for International Educational Research)
DME	Data Management Expert
EPSEM	Equal Probability of Selection Method
ESS	European Social Survey
ETS	Educational Testing Service
GPCM	Generalized partial credit model
GREG	Generalized regression estimator
HTML	Hypertext markup language
IAB	Institut für Arbeitsmarkt- und Berufsforschung (Institute for Employment Research)
IALS	International Adult Literacy Survey
ICT	Information and communication technologies
IDB	International Database
IEA DPC	International Association for the Evaluation of Educational Achievement – Data Processing and Research Center
IRT	Item Response Theory
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
ISIC	International Standard Industrial Classification of All Economic Activities
JK-1	Delete-one jackknife
JRA	Job Requirement Approach
LFS	Labour Force Survey
LRNR	Literacy-related nonrespondent
MOAS	Mother Of All Scripts
MOS	Measure of size
n/a	Not applicable
NRBA	Nonresponse bias analysis
NRW	Nonresponse weight
OECD	Organization for Economic Cooperation and Development
OLT	Open Language Tool
PBA	Paper-based assessment
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for International Student Assessment
PPS	Probability proportionate to size
PS-TRE	Problem solving in technology-rich environments
PSU	Primary sampling unit

ROA	Research Centre for Education and the Labour Market
RP	Response probability
SE	Standard error
SOEP	Sozio-oekonomisches Panel (Socio-economic Panel)
TAO	Testing Assisté par Ordinateur (computer-based testing)
UEW	Unknown eligibility weight
VM	Virtual machine
WZB	Wissenschaftszentrum Berlin für Sozialforschung (Berlin Social Science Center)
XLIFF	XML Localization Interchange File Format
XML	Extensible Markup Language
2PL	Two-parameter logistic model