



Open Access Repository

www.ssoar.info

Design and Key Features of the PIAAC Survey of Adults

Kirsch, Irwin; Yamamoto, Kentaro; Khorramdel, Lale

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Kirsch, I., Yamamoto, K., & Khorramdel, L. (2020). Design and Key Features of the PIAAC Survey of Adults. In D. B. Maehler, & B. Rammstedt (Eds.), *Large-Scale Cognitive Assessment: Analyzing PIAAC Data* (pp. 7-26). Cham: Springer. https://doi.org/10.1007/978-3-030-47515-4_2

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

Chapter 2

Design and Key Features of the PIAAC Survey of Adults



Irwin Kirsch, Kentaro Yamamoto, and Lale Khorramdel

Abstract This chapter gives an overview of the most important features of the Programme for the International Assessment of Adult Competencies (PIAAC) survey as it pertains to two main goals. First, only a well-designed survey will lead to accurate and comparable test scores across different countries and languages both within and across assessment cycles. Second, only an understanding of its complex survey design will lead to proper use of the PIAAC data in secondary analyses and meaningful interpretation of results by psychometricians, data analysts, scientists, and policymakers. The chapter begins with a brief introduction to the PIAAC survey followed by an overview of the background questionnaire and the cognitive measures. The cognitive measures are then compared to what was assessed in previous international adult surveys. Key features of the assessment design are discussed followed by a section describing what could be done to improve future PIAAC cycles.

2.1 Introduction

In today's world, what people know, and what they can do with this knowledge, matters more than ever—affecting both personal life outcomes and the well-being of societies. The demands of technologically infused economies, the rapid pace of change, and global competition have interacted to change the way we work and live. More and more, everyday tasks require the ability to navigate, critically analyse, and problem-solve in data-intensive, complex digital environments. Similarly, global forces have altered the workplace and increased the demand for more broadly skilled employees. Employers seek workers who can keep pace with rapidly changing

I. Kirsch (✉) · K. Yamamoto
Educational Testing Service, Princeton, NJ, USA
e-mail: ikirsch@ets.org

L. Khorramdel
National Board of Medical Examiners, Princeton, NJ, USA

© The Author(s) 2020
D. B. Maehler, B. Rammstedt (eds.), *Large-Scale Cognitive Assessment*,
Methodology of Educational Measurement and Assessment,
https://doi.org/10.1007/978-3-030-47515-4_2

technologies. As a result, they are looking for individuals who have skills that enable them to benefit from ongoing training programmes and, perhaps most importantly, have the ability and initiative to learn on their own and continuously upgrade what they know and can do. Claudia Goldin and Lawrence Katz (2008: 352) described the consequences of this new reality in their book *The Race between Education and Technology*:

As technological change races forward, demands for skills—some new and some old—are altered. If the workforce can rapidly make the adjustment, then economic growth is enhanced without greatly exacerbating inequality of economic outcomes. If, on the other hand, the skills that are currently demanded are produced slowly and if the workforce is less flexible in its skill set, then growth is slowed *and* inequality widens. Those who can make the adjustments as well as those who gain the new skills are rewarded. Others are left behind.

Recognising the ongoing changes that technology and globalisation are having on how we live and work, policymakers have become increasingly concerned not only about the levels of traditional literacy skills in their populations but also because of the growing importance of human capital and the broadening of the skills that will be needed to sustain productivity and social cohesion. The increased importance of human capital, and the learning associated with it, have led to a critical need for information about the distribution of knowledge, skills, and characteristics necessary for full participation in modern societies.

The Organisation for Economic Co-operation and Development (OECD) Programme for the International Assessment of Adult Competencies (PIAAC) took a significant step forwards in the assessment of adult skills by building on the pioneering work of two previous surveys implemented since the mid-1990s: the International Adult Literacy Survey (IALS, 1994–1998) and the Adult Literacy and Lifeskills Survey (ALL, 2003–2006).¹ As with the two earlier surveys, PIAAC was designed to provide internationally comparable data to help policymakers and other stakeholders better understand:

- The types and levels of adult skills that exist in each of the participating countries that are thought to underlie both personal and societal success
- The relationship between these skills and broader social and economic outcomes
- Factors that contribute to the development, maintenance, and loss of skills over the life cycle
- And help clarify some of the policy levers that could contribute to enhancing competencies

PIAAC has been planned by the OECD as an ongoing programme of work. The development and administration of the first cycle of PIAAC resulted in the largest

¹The **International Adult Literacy Survey (IALS)** was conducted between 1994 and 1998 as the first large-scale international comparative assessment designed to measure literacy skills of adults (ages 16–65 years old) in 22 countries and regions. Trend items from IALS were included in the Adult Literacy and Lifeskills Survey (ALL), conducted in 2003, and the Programme for the International Assessment of Adult Competencies (PIAAC), allowing data from IALS to be linked to trend data from participating countries in ALL and PIAAC.

and most innovative international survey of adults ever conducted. Administered in three rounds from 2012 through 2018 (i.e. at three different time points, with different countries being assessed at each time point), the first cycle of PIAAC was unprecedented in scope, assessing close to 200,000 adults across 38 countries. Twenty four countries completed and reported results in the first round, nine in the second, and five in the third.

As the first computer-based survey of its kind, PIAAC expanded what could be measured and changed how a large-scale assessment could be designed and implemented. These advances were the result of a number of key innovations, which included:

- Developing an integrated platform that handled computer-based instruments as well as paper-based instruments to allow the assessment of those adults who were unable or unwilling to take a computer-based test
- Designing and delivering items that mirrored the kinds of technology-based tasks increasingly required both in the workplace and everyday life
- Conducting a mode study that enabled continuity with, and links to, IALS and ALL
- Incorporating multistage computer-adaptive algorithms into a large-scale assessment to provide more reliable information about participants' skills and support a more complex assessment design
- Implementing automatically scored items across some 50 language versions of the cognitive instruments to improve scoring reliability and reduce the burden on participating countries
- Using process data, in particular timing information, to both enhance the interpretation of performance and evaluate the quality of the assessment data

2.2 What PIAAC Measures

As the first computer-based, large-scale adult literacy assessment, PIAAC reflects the changing nature of information, its role in society, and its impact on people's lives. While linked by design to IALS and ALL, including sets of questions from these previous surveys, PIAAC has refined and expanded the existing assessment domains and introduced two new domains as well. The main instruments in PIAAC included a background questionnaire and cognitive assessments focused on literacy, numeracy, reading components, and problem solving in technology-rich environments.²

²Reading components and problem solving were optional domains in Round 1. Of the countries that reported results in Round 1, most implemented the reading components assessment, with the exceptions being Finland, France, and Japan. Most implemented problem solving, with the exceptions being France, Italy, and Spain. In Rounds 2 and 3, no components were optional, with these two domains treated as core components.

2.2.1 *Background Questionnaire*

The PIAAC background questionnaire (BQ) was a significant component of the survey, taking up to one-third of the total survey time. The scope of the questionnaire reflects an important goal of adult surveys: to relate skills to a variety of demographic characteristics and explanatory variables. The information collected via the BQ adds to the interpretability of the assessment, enhancing the reporting of results to policymakers and other stakeholders. These data make it possible to investigate how the distribution of skills is associated with variables including educational attainment, gender, employment, and the immigration status of groups. A better understanding of how performance is related to social and educational outcomes enhances insight into factors related to the observed distribution of skills across populations as well as factors that mediate the acquisition or decline of those skills.

The BQ was the most detailed of its kind to date for a large-scale assessment of adults. Questions went well beyond age, gender, and job title. The questionnaire addressed issues such as skills used at work and home, focusing specifically on literacy, numeracy, and the use of digital technologies. Furthermore, it addressed learning strategies, civic engagement, and whether respondents had trust in government or other individuals. It also included a short section on a person's health and subjective well-being. The reader is referred to the following for more information about the comprehensiveness of the PIAAC BQ including a collection of publications using the PIAAC data (Maehler et al. 2020; OECD 2016a, b).

The questionnaire provided not only breadth but also depth in terms of its questions. Rather than simply asking a person's job title, it delved into the work involved. If, for example, a person worked in sales, questions were posed on whether he or she made presentations and how often. The questionnaire also asked whether he or she advised colleagues and had to work cooperatively.

Furthermore, it looked deeply into the kinds of literacy and numeracy skills used at home. Rather than simply asking how often a person used writing skills, for example, it asked whether the individual wrote letters, memos, or emails. It also asked about the individual's reading habits—whether the person read newspapers, magazines, or newsletters; whether he or she looked at professional journals; and so on. It also asked about use of a calculator for complex problems. Significantly, as PIAAC was the first large-scale assessment for adults developed as a computer-based assessment, the questionnaire also probed into information and communication technologies (ICT) skills used at work and at home, specifically asking questions about how often individuals used a computer, or the types of things they did with it, ranging from the types of programmes they used to whether their focus was on learning or socialising.

The questionnaire also included a Jobs Requirement Approach (JRA) section. The objective was to collect information on skills used at work in contrast to the demographic characteristics and other personal background information collected in the BQ (OECD 2013). This section was included because case studies have shown that skills beyond literacy—communication, teamwork, multitasking, and the ability

to work independently—are being rewarded in the labour market (Dickerson and Green 2004). The JRA was designed to assess the relevance of these skills.

One important new strategy with the questionnaire paid off with improved data on personal income. Income is chronically underreported in surveys (Pleis and Dahlhamer 2004), with rates of 20–50% of income having not been reported in the past (Moore et al. 2000). In PIAAC, categories were used that made respondents feel more comfortable to answer. The survey asked individuals to list income amounts they felt most comfortable sharing information about— annually, monthly, hourly, or by piece. Those unwilling to list a specific amount were asked whether they would provide amounts within specific ranges. With imputation techniques, it could be determined with some accuracy what those amounts were based on other variables such as occupation, industry, and age. PIAAC wound up with a total of 94.1% of respondents willing to report total earnings.

2.2.2 *Cognitive Domains*

The cognitive measures in PIAAC included literacy and numeracy, as well as the new domains of reading components and problem solving in technology-rich environments. The literacy and numeracy domains incorporated both new items developed for PIAAC and trend items from IALS and ALL. In order to maintain trend measurement, the PIAAC design required that 60% of literacy and numeracy items be taken from previous surveys, with the remaining 40% newly developed. In the case of literacy, items were included from both IALS and ALL. As numeracy was not a domain in IALS, all of the numeracy linking items came from ALL.

Like IALS and ALL, PIAAC included intact stimulus materials taken from a range of adult contexts, including the workplace, home, and community. As a computer-delivered assessment, PIAAC was able to include stimuli with interactive environments, such as webpages with hyperlinks, websites with multiple pages of information, and simulated email and spreadsheet applications.

To better reflect adult contexts as opposed to school-based environments, open-ended items have been included in international large-scale adult assessments since IALS. The innovation introduced in the first cycle of PIAAC was automatic scoring of these items, which contributed to improved scoring reliability within and across countries.

Literacy Literacy was defined in the first cycle of PIAAC as ‘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’ (OECD, 2012: 20). ‘Literacy’ in PIAAC does not include the ability to write or produce text—skills commonly falling within the definition of literacy. While literacy had been a focus of both the IALS and ALL surveys, PIAAC was the first to address literacy in digital environments. As a computer-based assessment, PIAAC included literacy tasks that required respondents to use electronic texts, including webpages, emails,

and discussion boards. These interactive stimulus materials included hypertext and multiple screens of information and simulated real-life literacy demands presented by digital media.

Reading Components The new domain of reading components was included in PIAAC to provide more detailed information about adults with limited literacy skills. Reading components represent the basic set of decoding skills that provide necessary preconditions for gaining meaning from written text. These include knowledge of vocabulary, ability to process meaning at the sentence level, as well as reading of short passages of text in terms of both speed and accuracy.

Adding this domain to PIAAC provided more information about the skills of individuals with low literacy proficiency than had been available from previous international assessments. This was an important cohort to assess, as it was known from previous assessments that there are varying percentages of adults across participating countries who demonstrate little, if any, literacy skills. Studies in the United States and Canada show that many of these adults have weak component skills, which are essential to the development of literacy and numeracy skills (Strucker et al. 2007; Grenier et al. 2008).

Numeracy The domain of numeracy remained largely unchanged between ALL and PIAAC. However, to better represent this broad, multifaceted construct, the definition of numeracy was coupled with a more detailed definition of numerate behaviour for PIAAC. Numerate behaviour involves managing a situation or solving a problem in a real context by responding to mathematical content, information, or ideas, represented in multiple ways (OECD 2012). Each aspect of numerate behaviour was further specified as follows:

- Real contexts including everyday life, work, society, and further learning.
- Responding to mathematical content, information, or ideas may require any of the following: identify, locate or access, act upon and use (to order, count, estimate, compute, measure, or model), interpret, evaluate or analyse, and communicate.
- Mathematical content, information, and ideas including quantity and number, dimension and shape, pattern, relationships and change, and data and chance.
- Representations possibly including objects and pictures, numbers and mathematical symbols, formulae, diagrams, maps, graphs and tables, texts, and technology-based displays.

Problem Solving in Technology-Rich Environments (PS-TRE) PS-TRE was a new domain introduced in PIAAC and represented the first attempt to assess it on a large scale and as a single dimension. While it has some relationship to problem solving as conceived in ALL, the emphasis in PIAAC was on assessing the skills required to solve information problems within the context of ICT rather than on analytic problems per se. PS-TRE was defined as ‘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 focuses on

Table 2.1 Domains assessed in PIAAC, ALL, and IALS

PIAAC	ALL (2003–2006)	IALS (1994–1998)
Literacy (combined prose and document)	Literacy (combined prose and document ^a)	Literacy (combined prose and document ^a)
	Prose literacy	Prose literacy
	Document literacy	Document literacy
Reading components		
Numeracy	Numeracy	
		Quantitative literacy
Problem solving in technology-rich environments		
	Problem solving	

Note.^aRescaled to form a single literacy scale combining the former separate prose and document literacy scales

the abilities to solve problems for personal, work and civic purposes by setting up appropriate goals and plans and accessing and making use of information through computers and computer networks’ (OECD 2012: 47).

The PS-TRE computer-based measures reflect a broadened view of literacy that includes skills and knowledge related to information and communication technologies—skills that are seen as increasingly essential components of human capital in the twenty-first century.

2.2.3 Relationship of PIAAC Domains to Previous Adult Surveys

As noted earlier, PIAAC was designed in a way that allowed for linking a subset of the domains assessed in the two earlier international surveys of adults—IALS and ALL. Table 2.1 shows the skill domains assessed in the three surveys. Shading indicates where the domains have been linked across the surveys.

IALS assessed three domains of literacy—prose, document, and quantitative. Prose literacy was defined as the knowledge and skills needed to understand and use *continuous* texts—information organised in sentence and paragraph formats. Document literacy represented the knowledge and skills needed to process documents or information organised in matrix structures (i.e. in rows and columns). The types of documents covered by this domain included tables, signs, indexes, lists, coupons, schedules, charts, graphs, maps, and forms. Quantitative literacy covered the skills needed to undertake arithmetic operations, such as addition, subtraction,

multiplication, or division either singly or in combination using numbers or quantities embedded in printed material.

The major change between IALS and ALL was the replacement of the assessment of quantitative literacy with that of numeracy and the introduction of the assessment of problem solving. Numeracy represented a broader domain than that of quantitative literacy, covering a wider range of quantitative skills and knowledge (not just computational operations) as well as a broader range of situations in which actors had to deal with mathematical information of different types, and not just situations involving numbers embedded in printed materials (Gal et al. 2005: 151). Problem solving was defined as ‘goal-directed thinking and action in situations for which no routine solution procedure is available’ (Statistics Canada and OECD 2005: 16).

In literacy, PIAAC differs from IALS and ALL in two main ways. First, literacy is reported on a single scale rather than on two separate (prose and document literacy) ones. For the purposes of comparison, the results of IALS and ALL were rescaled on the PIAAC literacy scale. Second, while the measurement framework for literacy in PIAAC draws heavily on those used in IALS and ALL, it expands the kinds of texts covered to include electronic and combined texts in addition to the continuous (prose) and noncontinuous (document) texts of the IALS and ALL frameworks. In addition, the assessment of literacy was extended to include a measure of reading component skills that was not included in previous assessments.

The domain of numeracy remains largely unchanged between ALL and PIAAC. PS-TRE constitutes a new domain. While it has some relationship to problem solving as conceived in ALL, the emphasis is on the skills necessary to solve ‘information problems’ and the solution of problems in digital environments rather than on analytic problem skills per se presented in paper-and-pencil format.

2.3 Assessment Design: Key Features

To provide accurate, valid, and stable measures of the domains and constructs described above, PIAAC is based on a complex survey or test design. There are two main features of this design. First, it is a matrix sampling design where a large item pool is administered to test-takers in a way that reduces the testing time for individuals while providing a broad construct coverage at the group and country level. More precisely, each test-taker responds only to a subset of items, but these subsets are linked throughout the design to enable the construction of a single joint scale for each domain. Second, the design is administered as a multistage adaptive test design (MST), which matches the administration of test items with regard to their difficulty to the proficiency level of test-takers. The first adaptive level directs a test-taker either to the paper- or the computer-based assessment branch based on his/her computer experience and skills. The second adaptive level directs test-takers to either more or less difficult items based on their responses to prior administered items. To enable the success of this complex design in the large variety of countries,

PIAAC implemented a field test prior to the main study to evaluate the developed instruments, the efficiency and implementation of the design, the data collection processes, and the computer-based testing platform.

In the following section, we will describe in more detail the different but related goals of the PIAAC field test and main study, give an overview of the advantages of implementing adaptive testing in PIAAC, and illustrate the core features of the final MST study design for the main study.

2.3.1 Field Test Versus Main Study

As stated previously, PIAAC is a cyclical cross-country survey that consists of a field test and a main study.

The *goal of the main study* was to provide policymakers, stakeholders, and researchers with data and test scores that are accurate and comparable across different countries and over time to enable fair and meaningful comparisons as well as a stable measure of trends. To achieve this goal, PIAAC implemented an MST design that allows for higher test efficiency and more accurate measurements within the specified testing time, especially at the extreme ends of the proficiency scale. Moreover, the design needed to provide a successful link across the different cognitive domains within the PIAAC assessment cycle and between PIAAC and prior adult surveys (IALS and ALL). The main study design will be illustrated in more detail in Sect. 2.3.2. To ensure that all goals were met, a field test was implemented.

The *goal of the field test* was to prepare for the main study instrument, the MST design, computer delivery platform, data collection, and analysis. With regard to the MST design, the field test was used to examine the role of test-takers' computer familiarity, evaluate the equivalence of item parameters between the paper-based assessment (PBA) and computer-based assessment (CBA), and establish initial item parameters based on item response theory (IRT) models. The item parameters were used to select items for the final PIAAC instruments and construct the adaptive testing algorithm for branching test-takers in the final MST design. More details about the PIAAC field test design and analysis in preparation of the final PIAAC MST design can be found in the PIAAC Technical Report (OECD 2013; Kirsch and Yamamoto 2013).

2.3.1.1 Advantages and Efficiency of Multistage Testing in PIAAC

PIAAC was one of the first international large-scale assessments to introduce an adaptive test design in the form of MST. Using an MST design allowed PIAAC to assess a broader range of proficiency levels more accurately within and across countries. This is important given that more and more countries are participating in this international large-scale survey.

MST increases the efficiency, validity, and accuracy of the measured constructs by matching the administration of test items to the proficiency level of test-takers. This leads to an improvement of proficiency estimation and a reduction in measurement error across the entire proficiency distribution (Lord 1980; Wainer 1990) and particularly with regard to the ends of the proficiency scale (Hambleton and Swaminathan 1985; Lord 1980; Weiss 1974). A reduction of the linking error (Wu 2010) and a potential increase of test-taker engagement (Arvey et al. 1990; Asseburg and Frey 2013), especially for low-performing respondents (Betz and Weiss 1976), are additional advantages.

MST is an extension of item-level adaptive testing that allows the choice of the next item set as opposed to the selection of single items. Since international large-scale assessments make use of item sets in the test design, the implementation of MST for introducing adaptive testing is a reasonable choice. In item sets (or units), several items share the same stimulus. In PIAAC, item sets are used as intact entities (i.e. are not split), which provides the ability to control the presentation of items across different test forms for better construct coverage and balancing item position to prevent bias on parameter estimation. Moreover, MST accumulates more information after each adaptive step compared to approaches that use single items for each adaptive decision or path. This can lead to greater accuracy in the decision of the next adaptive path and reduce the likely dependency of the adaptive selection on item-by-country interactions (Kirsch and Thorn 2013). More details about benefits of MST for international large-scale assessments can be found in Yamamoto et al. (2018). In summary, the MST approach in PIAAC allows for matching item difficulty with the abilities of test-takers while meeting other design requirements (item parameter estimation, broad construct coverage, balancing item content, item type, and the position of items, linking) at the same time.

The PIAAC MST design was able to achieve its main goal—improvement in measurement precision—especially for higher and lower proficiency levels. Based on the international common item parameters of PIAAC, the MST design was 10–30% more efficient for literacy and 4–31% more efficient for numeracy compared to the nonadaptive average linear tests of equal length. In other words, it is possible to obtain the same amount of test information as one might expect from a test that is 10–30% longer with regard to literacy and 4–31% longer with regard to numeracy. There was no proficiency range where MST was less informative, with more gains for extreme scale scores.

2.3.2 *Main Study Design*

PIAAC used a variant of matrix sampling where each test-taker was administered a subset of items from the total item pool. Hence, different groups of test-takers answered different sets of items, leading to missing data by design. PIAAC consisted of a BQ administered at the beginning of the survey (30–40 min) followed by a cognitive *assessment* (60 min) measuring the four domains literacy, numeracy,

Table 2.2 Terminologies for describing the PIAAC main study design

PBA (nonadaptive)	CBA (adaptive)
<i>Item:</i> Refers to a task to which an examinee is directed to provide a response. The response is coded based on a coding guide; in PIAAC, all items are machine coded	<i>Item:</i> Same as PBA
<i>Unit:</i> Refers to a short, mutually exclusive set of items in the PIAAC adaptive test design	<i>Unit:</i> Same as PBA
<i>Cluster:</i> Refers to a mutually exclusive set of items in the PBA; one cluster takes 30-min testing time on average	<i>Block:</i> A set of units in the PIAAC adaptive test design; each respondent receives two blocks: one in adaptive Stage 1 and one in adaptive Stage 2
<i>Booklet:</i> Each respondent in the nonadaptive PBA receives one booklet; a booklet consists of two 30-min clusters (60 min on average)	<i>Module:</i> Refers to a domain-specific set of two blocks across the adaptive stages in the PIAAC adaptive test design (one Stage 1 block and one Stage 2 block); one module takes 30-min testing time on average; each examinee receives two cognitive domains, i.e. two modules (60 min on average)

Notes. PBA paper-based assessment, CBA computer-based assessment

reading components (RC), and problem solving (PS-TRE). Furthermore, a *link to prior adult surveys* (IALS and ALL) was established through 60% of literacy and numeracy linking items that are common across the different surveys. The different item types in PIAAC in the CBA (highlighting, clicking, single choice, multiple choice, and numeric entry) were scored automatically and instantaneously by the computer-based platform based on international and national scoring rules. This was done to enable adaptive testing in the CBA. In the following, we describe the PIAAC main study design in detail using the terminologies described in Table 2.2.

2.3.2.1 Levels of Adaptiveness

The PIAAC MST design as displayed in Fig. 2.1 was *adaptive on different levels*. The *first level* of adaptiveness accounted for test-takers' computer familiarity. Test-takers were either routed to the PBA or the CBA based on their responses to questions from the BQ and a core set of questions focusing on ICT skills. Test-takers who reported no familiarity with computers were routed to the PBA, as were those who refused to take the test on the computer. Test-takers who reported familiarity with computers in the main study were routed to the CBA. The *second level* of adaptation was within the CBA cognitive assessment. PIAAC used a probability-based multistage adaptive algorithm, where the cognitive items for literacy and

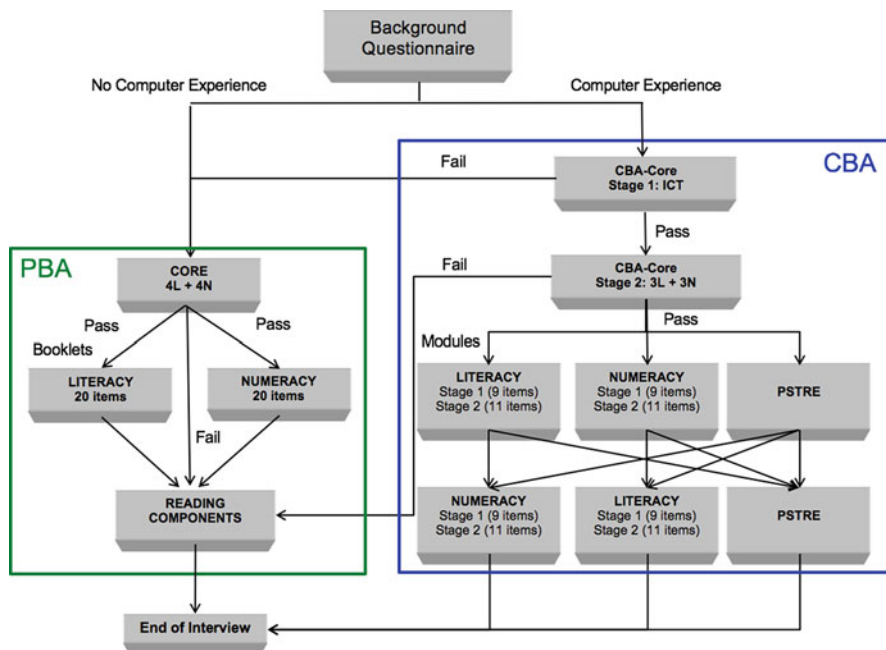


Fig. 2.1 PIAAC MST main study design

numeracy were administered to test-takers in an adaptive way. In other words, more able test-takers received a more difficult set of items than less able respondents did. Note that PS-TRE was not administered adaptively.

2.3.2.2 PBA and CBA Branches

The *PBA branch* started with a 10-min core assessment of literacy and numeracy. Test-takers who performed at or above a minimum standard were randomly assigned to a 30-minute cluster of literacy or numeracy items, followed by a 20-min assessment of reading components. The small proportion of test-takers who performed poorly on the PBA core items did not receive literacy and numeracy items and were routed directly to the reading component items.

The *CBA branch* started with the CBA core section, which was composed of two stages taking approximately 5 min each. Poor performance on either stage of the CBA core sections resulted in switching over to the appropriate sections of the PBA instruments. Test-takers who failed CBA Core Stage 1 (which contained ICT-related items) were redirected to the PBA. Those who passed CBA Core Stage 1 but failed CBA Core Stage 2 (which contained six cognitive items) were administered only the reading component items. Those who performed well on both CBA core

Table 2.3 Design of the main study CBA instruments for literacy and numeracy in the integrated design

Stage 1 (18 unique items—9 items per block. Each respondent takes 1 block)							
	<i>Unit A1</i> 4 items	<i>Unit B1</i> 5 items	<i>Unit C1</i> 4 items	<i>Unit D1</i> 5 items			
Block 1-1	X	X					
Block 1-2		X	X				
Block 1-3			X	X			
Stage 2 (31 unique items—11 items per block. Each respondent takes 1 block)							
	<i>Unit A2</i> 6 items	<i>Unit B2</i> 5 items	<i>Block C2</i> 3 items	<i>Unit D2</i> 3 items	<i>Unit E2</i> 3 items	<i>Unit F2</i> 5 items	<i>Unit G2</i> 6 items
Block 2-1	X	X					
Block 2-2		X	X	X			
Block 2-3				X	X	X	
Block 2-4						X	X

Note. One block consists of two or three item units; one module within a stage consists of two blocks

sections were routed to one of three possible CBA module combinations (each taking approximately 50 min):

1. A combination of literacy and numeracy modules
2. A PS-TRE module combined with either a literacy or a numeracy module
3. Only PS-TRE modules

The *literacy* and *numeracy modules* each consisted of two adaptive stages. Each stage contained a number of blocks varying in difficulty, with each block consisting of several item units (a unit is a mutually exclusive set of items). In each stage, only one block was delivered to a test-taker. The blocks within one stage were linked through a common item unit (see Table 2.3) to provide stable item parameter estimates in the main study. Within each of these modules, a test-taker took 20 items (9 in Stage 1; 11 in Stage 2). Hence, test-takers receiving literacy in Module 1 and numeracy in Module 2 (or vice versa) answered 40 items. Each module was designed to take an average of 30 min. The *PS-TRE modules* were not adaptive and comprised seven items in Module 1 and seven items in Module 2. The PS-TRE modules were also designed to take an average of 30 min. Table 2.3 provides an overview of the design of the MST Stages 1 and 2.

2.3.2.3 Controlled Item Exposure Rates and Module Selection

The diversity of countries, languages, and educational backgrounds would likely have resulted in certain subpopulations being exposed to only a small percentage of items when using a deterministic assignment of stages. This could have reduced the content coverage for single cognitive domains per country and the comparability of the PIAAC survey across countries. For achieving comparable data and test scores, a

Table 2.4 Number of cognitive items per assessment mode and domain in PIAAC

Domain	Assessment mode	Number of items
Literacy	CBA	52
	PBA	24
Numeracy	CBA	52
	PBA	24
PS-TRE	CBA	14
Reading components	PBA	100

Note. 18 literacy and 17 numeracy items were linking items between the PBA and CBA assessment mode, meaning they were identical. Thus, PIAAC contained a total of 131 unique items (excluding reading components)

set of *conditional probability tables* was used to control the *item exposure rates* for specified subpopulations (Chen et al. 2014). For more information on the module selection based on conditional probabilities, and for practical examples, see the PIAAC Technical Report (OECD 2013) as well as Yamamoto et al. (2018).

2.3.2.4 Items and Comparability

The PIAAC MST design was based on 76 literacy and 76 numeracy items that were scored dichotomously and 14 PS-TRE items that were scored dichotomously or polytomously. Table 2.4 provides an overview of the number of items per assessment mode (PBA and CBA).

Item position effects at the cross-country level as well as the *comparability of item parameters* across countries (item-by-country interactions) were examined in the field test and main study (OECD 2013; Yamamoto et al. 2018). There was the possibility that results would show a slight cluster position effect for literacy modules (2.9%) and numeracy modules (1.2%) on the per cent of correct responses. However, the IRT scaling provided comparable item parameters achieving high comparability and measurement invariance (92% and 94% for literacy and 93% and 97% for numeracy in the PIAAC Round 1 and Round 2 assessments, respectively). Overall, item parameters were shown to be stable and comparable across the different countries and languages.

2.4 Sampling Requirements

The target population for PIAAC included adults between the age of 16 and 65 years, excluding adults in institutions (e.g. prisons). The sampling unit for PIAAC was individuals or, in the case of countries not having register-based sampling frames, the household. In the latter case, each sampled household was

administered a screener to determine the eligibility of household members. Within households, each selected adult was administered the BQ and cognitive assessment.

Countries also had national options to include oversamples of key subpopulations or to include additional subpopulations in their PIAAC target population (e.g. adults aged 66 to 74 years). Therefore, the sampling plan included guidelines for the national options chosen by countries as well as specifications for any necessary augmentation of the sample size to accommodate the analysis requirements for these additional subsamples.

The core sample design was a stratified multistage clustered area sample. However, deviations from the core design were expected due to geographically small countries that have less clustering and fewer stages of sampling. Some countries had lists of households or persons already available from population registries. The general approach was to allow for flexibility in the sample design, conduct a thorough assessment of the quality of sampling frames, and prepare to adapt to each country's best sampling scenario.

The minimum sample size required to produce reliable estimates of skills at the national level in a country was between $N = 4000$ and $N = 5000$. As stated above, all countries had the option of boosting sample size and oversampling to obtain estimates for subpopulations of special interest or to increase sample size to get reliable estimates at the subnational level (e.g. states, regions, or provinces or language groups). As the field test had distinct purposes that differed from those of the main study, their sampling requirements also differed. Since the field test was not used for any reporting, and was designed solely to test operational issues along with instrument quality, fewer respondents were needed. For example, only 1500 completed cases were required in PIAAC. The reader is referred to the PIAAC Technical Report for more detailed information on sampling requirements (OECD 2013).

2.5 Future Cycles of PIAAC: Potential Improvements

While many of the innovations from the first cycle of PIAAC are being carried forward to the second cycle, ongoing technological developments are expected to enable the implementation of new innovations that will be explored to further improve the accuracy and comparability of the data and the measurement of trend. They include:

- *New constructs:* New types of interactive stimulus materials and item formats can be incorporated to extend what is measured. In addition to measuring reading component skills, the component skills will be extended to the numeracy domain. Moreover, a new domain—adaptive problem solving—will replace PS-TRE.
- *Existing constructs and linking:* The current number of items for literacy and numeracy will be increased to provide better overall construct coverage and measurement along each scale. Furthermore, the number of core items will

be doubled to provide better measurement of low-performing adults in each participating country while not requiring that they take the full assessment. The measures of literacy and numeracy will be linked between PIAAC Cycle 1 and Cycle 2 as well as to previous adult surveys (IALS, ALL).

- *Process data and adaptive algorithm:* The use of process information from computer-based tests, such as timing data, will be explored to refine the adaptive algorithms for multistage adaptive testing to increase both the validity and efficiency of adaptive testing.
- *Delivery mode and hardware:* The use of tablet devices will be explored for possibly replacing the paper-based assessment. The tablet devices will need to be of high quality to ensure that the touch sensitivity is sufficiently responsive to user input. The tablet will be connected to a keyboard for the interviewer (for administering the BQ) and to a stylus for the test-taker (for completing the cognitive assessment). Another possibility would be to allow respondents to complete the BQ on a tablet rather than having it administered by the interviewer. The stylus should allow the tablet to function much like a paper-and-pencil instrument in terms of not requiring much ICT skill without compromising the overall functionality and item types that are feasible on a technology platform. Increasing the number of test-takers for the CBA by using tablets would reduce the need for scoring paper-based responses (which improves scoring reliability), more participants would be able to benefit from the MST design, and more would be able to take the newly developed innovative items that are administered only in the CBA. However, different test designs will be available, especially for countries that are not able to switch to tablet devices and for test-takers who could not or chose not to use the tablet or laptop. For any of these options, studies would need to be conducted to learn more about the feasibility and impact of using alternative devices before they can be incorporated into the Cycle 2 main study. Device effects are an important consideration for trend items from literacy and numeracy with regard to the comparability of Cycle 1 and Cycle 2 and to test-takers with limited technology skills.
- *New software for data capture:* The use of new technologies for capturing oral proficiencies of test-takers with limited literacy skills could be explored. Albeit, this would require a good deal more research and development. Work is being done around spoken language tests that are automatically delivered and scored (see, e.g. Bernstein et al. 2010), and it could be explored whether comparable measures could be developed across languages for PIAAC.
- *Accessibility:* XML and web-based technologies will be used to develop data products and analysis systems that can accommodate a constantly expanding set of analysis, visualisation, and reporting tools to make the PIAAC data more accessible and powerful for a range of users (e.g. test-takers with certain disabilities).

The second cycle of PIAAC will need to balance innovation with the ongoing constraints of this survey. These include the importance of maintaining trend measurement and a recognition that the testing population includes individuals

who range broadly in terms of both age and familiarity with technology, as well as educational backgrounds and proficiencies. All possible improvements and innovations introduced in a second cycle of PIAAC could have considerable impact on the test design and will need to be considered when analysing the future PIAAC data.

2.6 Summary and Outlook

PIAAC needs to meet the goals and standards of international large-scale surveys while, at the same time, dealing with certain constraints and challenges. The major goal of PIAAC is to provide comparable, accurate, and fair measures of literacy, numeracy, problem solving, and reading component skills across participating countries, groups within countries, and different PIAAC cycles and prior adult surveys (i.e. across time points) to provide a stable measurement of trends. One important challenge PIAAC faces is the variability of proficiencies across and even within countries, as test-takers with a broad range in age (16–65 years) and educational levels are tested in multiple languages often associated with diverse cultural backgrounds. The PIAAC test design was developed to account for these constraints. The heart of the design is MST, which better matches the administration of test items to the proficiency level of test-takers. This provides an overall increase in test efficiency and accuracy within and across countries. The design also helps reduce the possible impact of item position and mode effects as well as item-by-country (and item-by-language) interactions. The improved measurement also allows for establishing a stable link over time and across assessment modes and different countries and languages.³

The PIAAC MST design uses information from both the BQ and cognitive assessment and was based on two levels of adaptation: (1) based on test-takers' computer skills and experience, they were routed to either PBA or CBA and (2) within the CBA, test-takers' proficiency levels with regard to responses to prior cognitive items as well as information about their educational level and native language were used to assign the different adaptive stages. A probability-based multistage adaptive algorithm was used to control the item exposure rate to enable a broad construct coverage and minimise item-by-country interactions.

³Please see the PIAAC Technical Report (OECD 2013) for more details around the development, implementation and analysis of the survey.

2.6.1 *What to Keep in Mind When Using PIAAC Data for Analysis*

The use of the data resulting from this complex test design for secondary analysis requires a good understanding of the design features. In the following, we summarise some of the most important points which should be considered when analysing the data.

- *Plausible values*: For secondary analysis (i.e. analysis based on the final test scores provided in the public use data file), plausible values should be used instead of raw responses, as they account for uncertainty in the measurement and reduce measurement error. Moreover, plausible values are placed on a common scale that allows for comparing different subgroups and countries in a fair and meaningful way. For details about the use of plausible values in analysis, see Chap. 3 in this volume.
- *Missing values*: PIAAC is based on an incomplete balanced block design. This means that every test-taker responded to just a subset of items, and the data include missing values. However, all items are linked together and can be placed on a common scale. In addition to these missing values by design, there are other types of missing data such as omitted responses (an item was presented, but the test-taker chose not to respond) and not-reached items. More information on different types of missing values can be found in the PIAAC Technical Report (OECD 2013). It is strongly recommended to use the plausible values for secondary analysis. However, if analysing the raw responses is needed, researchers and analysts have to consider how to treat these different types of missing values; again, the PIAAC Technical Report provides guidance in this regard.
- *Different administration modes due to adaptive testing*: PIAAC is based on an MST design. This means that some test-takers took PIAAC on paper, while the majority took it on computer.
- *Different domains due to administration mode and adaptive testing*: Not all test-takers received all cognitive domains. All test-takers responded to literacy and numeracy items, but not all received reading component or problem-solving (PS-TRE) items. All test-takers who received the PBA responded to reading component items (but not to PS-TRE items). Test-takers who received the CBA responded to problem-solving items; only a subset of test-takers from the CBA received reading component items.

PIAAC is the largest and most innovative assessment of adults in the world. It is both linked to and builds on two earlier adult surveys that allows for the measurement of changes in the distributions of adult skills among countries that have participated in all surveys. It also builds on the work of these two earlier surveys by assessing the use of digital texts and skills that better reflect the ways in which adults now access, use, and communicate information.

As reflected in the wide range of publications and papers that have been developed, when used properly and in a thoughtful way, the PIAAC dataset can provide policymakers, stakeholders, and researchers with a rich and accurate source of information to better understand the distributions of human capital in their country and the connections between these skills and important social, educational, and labour market outcomes. The next chapter will cover the statistical background of the PIAAC dataset. More precisely, Chap. 3 will illustrate the computation and correct use of plausible values, which are multiple imputations of group-level test scores for calculating group-level statistics in secondary analysis.

References

- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, *43*, 695–716. <https://doi.org/10.1111/j.1744-6570.1990.tb00679.x>.
- Asseburg, R., & Frey, A. (2013). Too hard, too easy, or just right? The relationship between effort or boredom and ability-difficulty fit. *Psychological Test and Assessment Modeling*, *55*, 92–104.
- Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, *27*(3), 355–377.
- Betz, N. E., & Weiss, D. J. (1976). *Psychological effects of immediate knowledge of results and adaptive ability testing* (Research Report No. 76-4). Minneapolis: University of Minnesota, Psychometric Methods Program.
- Chen, H., Yamamoto, K., & von Davier, M. (2014). Controlling MST exposure rates in international large-scale assessments. In D. Yan, A. A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 391–409). Boca Raton: Chapman and Hall/CRC.
- Dickerson, A., & Green, F. (2004). The growth and valuation of computing and other generic skills. *Oxford Economic Papers-New Series*, *56*, 371–406.
- Gal, I., van Groenestijn, M., Manly, M., Schmitt, M. J., & Tout, D. (2005). Adult numeracy and its assessment in the ALL Survey: A conceptual framework and pilot results. In S. Murray, Y. Clermont, & M. Binkley (Eds.), *Measuring adult literacy and life skills: New frameworks for assessment* (Catalogue 89-552-MIE, No. 13). Ottawa: Statistics Canada.
- Goldin, C., & Katz, L. F. (2008). *The race between education and technology*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Grenier, S., Jones, S., Strucker, J., Murray, T. S., Gervais, G., & Brink, S. (2008). *International adult literacy survey: Learning literacy in Canada: Evidence from the international survey of Reading skills*. Ottawa: Statistics Canada.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.
- Kirsch, I., & Thorn, W. (2013). The Programme for International Assessment of Adult Competencies—An Overview. In Organisation for Economic Co-operation and Development (2013), *Technical report of the Survey of Adult Skills (PIAAC)*, foreword (pp. 5–24), PIAAC, OECD Publishing. Retrieved from <http://www.oecd.org/site/piaac/All%20PIACC%20Technical%20Report%20final.pdf>
- Kirsch, I., & Yamamoto, K. (2013). PIAAC assessment design. In Organisation for Economic Co-operation and Development (2013), *Technical report of the Survey of Adult Skills (PIAAC)*, (pp. 27–43). Paris: OECD Publishing. Retrieved from <http://www.oecd.org/site/piaac/All%20PIACC%20Technical%20Report%20final.pdf>

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Lawrence Erlbaum Associates.
- Maehler, D. B., Jakowatz, S., & Konradt, I. (2020). *PIAAC bibliographie 2008 – 2019*. GESIS Papers, Nr. 04/2020. Cologne: GESIS – Leibniz Institute for the Social Sciences. <https://doi.org/10.21241/ssoar.67732>.
- Moore, J., Stinson, L. L., & Welniak, E. J., Jr. (2000). Income measurement error in surveys. *Journal of Official Statistics*, 16(4), 331–361.
- Organisation for Economic Co-operation and Development (OECD). (2012). *Literacy, numeracy and problem solving in technology-rich environments: Framework for the OECD Survey of Adult Skills*. <https://doi.org/10.1787/9789264128859-en>
- Organisation for Economic Co-operation and Development (OECD). (2013). *Technical report of the Survey of Adult Skills (PIAAC)*, (pp. 406–438). Retrieved from http://www.oecd.org/site/piaac/Technical%20Report_17OCT13.pdf
- Organisation for Economic Co-operation and Development (OECD). (2016a). *Skills matter: Further results from the Survey of Adult Skills studies*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development (OECD). (2016b). *The Survey of Adult Skills: Reader's companion* (2nd ed.). Paris: OECD Publishing.
- Pleis, J. R., & Dahlhamer, J. M. (2004). Family income response patterns for varying levels of income detail: An analysis of the National Health Interview Survey (NHIS). In *Proceedings of the American Statistical Association, Section on survey research methods* (CD-ROM) (pp. 4200–4207). Alexandria: American Statistical Association.
- Statistics Canada and Organisation for Economic Co-operation and Development (OECD). (2005). *Learning a living: First results of the adult literacy and life skills survey*. Ottawa/Paris: Authors.
- Strucker, J., Yamamoto, K., & Kirsch, I. (2007). *The relationship of the component skills of reading to IALS performance: Tipping points and five classes of adult literacy learners*. Cambridge, MA: Harvard Graduate School of Education, National Center for the Study of Adult Learning and Literacy.
- Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale: Lawrence Erlbaum Associates.
- Weiss, D. J. (1974). *Strategies of adaptive ability measurement* (Research Report No. 74-5). Minneapolis: University of Minnesota, Psychometric Methods Program.
- Wu, M. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice*, 29, 15–27. <https://doi.org/10.1111/j.1745-3992.2010.00190.x>.
- Yamamoto, K., Khorramdel, L., & Shin, H. J. (2018). Multistage adaptive testing design in PIAAC. *Psychological Test and Assessment Modeling*, 60, 347–368.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

