

SHORT REPORT

Open Access



# PIAAC: a new design for a new era

Irwin Kirsch\* and Mary Louise Lennon\*

\*Correspondence:  
ikirsch@ets.org; mlennon@ets.org  
Educational Testing  
Service, 660 Rosedale Road,  
Princeton, NJ 08541, USA

## Abstract

As the largest and most innovative international assessment of adults, PIAAC marks an inflection point in the evolution of large-scale comparative assessments. PIAAC grew from the foundation laid by surveys that preceded it, and introduced innovations that have shifted the way we conceive and implement large-scale assessments. As the first fully computer-delivered survey of adults, those innovations included: a comprehensive assessment design involving multistage adaptive testing; development of an open-source platform capable of delivering both cognitive measures and nationally-specific background questionnaires; automated scoring of open-ended items across more than 50 languages; enhanced cognitive measures that included electronic texts and interactive stimuli; the inclusion of new item types and response modes; and the use of log file and process data to interpret results. This paper discusses each of these innovations along with the development of data products and dissemination activities that have extended the utility of the survey, providing today's policy makers with information about the extent to which adults possess the critical skills required for both their own success and the health and vibrancy of societies around the world. As this paper suggests, the innovations introduced via PIAAC broadened the relevance and utility of the survey along with the accuracy and validity of the data, strengthening the foundation upon which future surveys can continue to build.

**Keywords:** PIAAC, Large-scale assessment history, Computer-based assessment, Assessment design, Multistage adaptive testing, Literacy, Numeracy, Reading components

## Background

In the early 1980s, Samuel Messick, Albert Beaton and Frederick Lord, along with others from Educational Testing Service (ETS), proposed a new design for the U.S. National Assessment of Educational Progress, or NAEP, motivated by a changing educational, social and political landscape. Calling this a “new design for a new era”, their reconceptualization of this large-scale assessment presented a conceptual framework that was *innovative* in its psychometric methodology, *comprehensive* in its impact on processes and procedures ranging from sampling to instrument development, data collection, analysis and dissemination, and *protective* of continuity in that it maintained and enhanced the examination of trends (Messick et al. 1983).

Some thirty years later, the Programme for the International Assessment of Adult Competencies (PIAAC), the latest in a series of large-scale international assessments focusing on adult populations, required a new design for yet another new era. Today's era, one characterized by technological innovation and increasing globalization,

demands that adults develop new types and levels of skills to meet rapidly changing work conditions and societal demands. As a result, a new design, built on the foundation laid by NAEP and other large-scale assessments, was needed to create PIAAC—a new assessment that could provide policy makers and other stakeholders with profiles of adults both within and across countries in terms of the knowledge, skills and competencies that are thought to underlie both personal and societal success.

Initiated by the Organisation for Economic Co-operation and Development (OECD), and first conducted in 2012, PIAAC profiles the knowledge, skills and competencies of adults ages 16–65. Administered in three rounds from 2012 through 2019, PIAAC is unprecedented in scope—assessing close to 200,000 adults across 38 countries.<sup>1</sup> As the first computer-based survey of its kind, PIAAC expanded both *what* could be measured and *how* a large-scale assessment could be designed, implemented, and administered to respondents in participating countries. Innovations in the first cycle included:

- Developing a platform capable of reliably delivering both the cognitive instruments and nationally-specific versions of the background questionnaire in household settings, capturing and exporting all respondent data—and doing so with no data loss;
- Developing an integrated assessment design that included both computer- and paper-based instruments;
- Designing and delivering items that mirrored the kinds of technology-based tasks increasingly required in both the workplace and everyday life;
- Implementing items capable of being automatically scored across some 50 language versions of the cognitive instruments;
- Incorporating multistage computer-adaptive algorithms into a large-scale assessment; and
- Using process data, in particular timing information, to both enhance the interpretation of performance and evaluate the quality of the assessment data.

Like the new design conceived and realized for NAEP, PIAAC's design and implementation was innovative, comprehensive in scope, and protective of the foundation laid by earlier international assessments of adults.

### **A brief history of large-scale assessments**

Broadly defined, large-scale assessments are surveys of knowledge, skills, or behaviors in one or more given domains. The goal of large-scale assessments is to describe populations of interest. Consequently, these assessments focus on group scores, in contrast to testing programs that focus on the assessment of individuals. The impetus for large-scale assessments has always been some call for the collection of comparable information about the skills possessed by a population in order to better understand how those skills are related to educational, economic and social outcomes (Kirsch et al. 2013). As

---

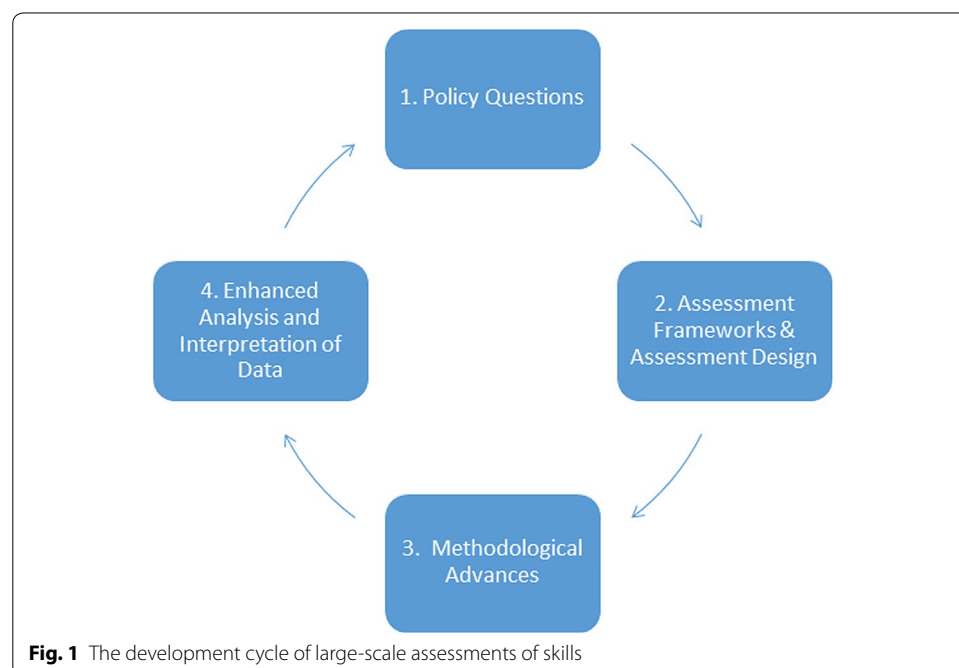
<sup>1</sup> Participating countries include: Round 1 (24 countries)—Australia, Austria, Belgium (Flanders), Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Netherlands, Norway, Poland, Republic of Korea, Russian Federation (results reported separately due to data problems), Slovak Republic, Spain, Sweden, United Kingdom (England and Northern Ireland), United States; Round 2 (9 countries)—Chile, Greece, Indonesia, Israel, Lithuania, New Zealand, Singapore, Slovenia, Turkey; Round 3 (5 countries)—Ecuador, Hungary, Kazakhstan, Mexico, Peru.

a result, the development of large-scale assessments follows a four-step cycle, as illustrated in Fig. 1 below.

The conceptualization of each assessment is motivated by a set of policy questions. Such questions lead to decisions around what should be measured and which populations should be assessed. For example, as globalization and technology continue to impact economies and everyday life, we have seen a growing interest among policy makers in comparative international surveys of both student and adult populations that focus on the types and level of skills needed for success in school or adult life. This interest has led not only to a broadening of what skills should be assessed but also who should be tested.

As the second step in the process, assessment frameworks and designs must be optimized to best address the policy questions of interest. Expanding what is measured to include new types and levels of skills requires new frameworks that define innovative domains as well as new designs to implement increasingly complex assessments. Operationalizing these assessments then requires new methodologies. For example, enhanced translation and verification methodologies are needed to ensure the comparability of assessment instruments across the range of languages and cultures participating in international surveys. Similarly, new applications of computer-based testing are needed to more accurately reflect the ways in which students and adults access, use and communicate information. As the final step in the cycle, the extended types and amount of data that result from new and innovative assessments drive the need for enhanced data analysis tools and methodologies, resulting in richer interpretative schemes.

The important point about this cycle is that it is the increasingly complex questions being asked by policy makers that drive the expansion of what should be assessed, the designs and methodologies used to develop and implement these surveys, and the



analysis and interpretation of the survey data. The growing importance of education and skills and the resulting need to better understand how attainment and skills are distributed both within and across countries lead, in turn, to new questions that form the basis of future assessment cycles. This cycle of inquiry, conceptualization, implementation and interpretation has played out in both national and international large-scale assessments since the mid-twentieth century.

### **Assessing student populations**

Large-scale assessments that compare the skills and knowledge demonstrated by populations across countries are relatively recent endeavors. This work began in the late 1950s, with a study designed to investigate the feasibility of developing and conducting an assessment of 13-year-olds in 12 countries. Known as the Pilot Twelve-Country Study, this assessment of academic skills and non-verbal ability was conducted from 1959-to 1962 by the International Association for the Evaluation of Educational Achievement (IEA). Participating countries included Belgium, England, Finland, France, Germany, Israel, Poland, Scotland, Sweden, Switzerland, the United States and Yugoslavia. Beyond the skills data that was collected, the critical finding from this pioneering effort was that it was possible to construct common instruments that worked in a comparable manner across different cultures and languages (Naemi et al. 2013). This knowledge laid the foundation for future large-scale international assessments of both students and adults.

Around this same period, in response to concerns about the lack of systematic data on the educational attainment of students in the United States, a number of prominent scholars and policy makers developed a plan for a periodic national assessment of student learning. The result was NAEP, which conducted its first assessment of in-school 17-year-olds in 1969.

The new design for NAEP, as introduced above, was proposed in 1983 and driven by policy questions focused on a desire to better understand how student competencies related to national concerns, human resource needs, and school effectiveness. Messick and his colleagues (Messick et al. 1983) argued that NAEP data should be relevant to questions across all three areas, including:

- Were students learning the skills necessary to ensure an educated populace that would contribute to the nation's political and economic well-being in the 1980s and beyond?
- Were students being equally well prepared, regardless of where they lived or their ethnicity, economic circumstances and social standing?
- Were students being prepared to meet the evolving work force needs of the country?
- How did particular curricula and teaching methods relate to achievement?

In order to answer such questions, NAEP needed to move beyond its original design and methodologies, which yielded interpretations of the data that were fixed to the individual items used in the assessments. The framework put forth by Messick and his colleagues for the new NAEP design addressed the desire to move beyond such limited interpretations and, in doing so, changed the face of large-scale assessments.

Central to the new design was a proposal to employ item response theory (IRT), which, they argued, had important advantages compared to the classical methods used previously in that it directly supports the creation of comparable scales across multiple forms of a test. In addition to incorporating IRT-based methodology, the work on NAEP led to the development of additional methodologies including marginal estimation procedures that could optimize the reporting of proficiency scales based on very complex designs (von Davier et al. 2006). The introduction of balanced incomplete block (BIB) spiraling, where each student is administered only a small subset of the total item pool, was another important innovation as that made it possible to maximize the coverage of the assessment constructs while reducing the burden on the individual test taker. Taken together, the application of these new psychometric methodologies enriched the body of information that these assessments could provide to policy makers.

The evolution of NAEP thus resulted in the development, application and refinement of innovative psychometric methodologies that have been used and expanded in subsequent international large-scale assessments. Examples are studies of student skills conducted by the IEA, including the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS), as well as the OECD's Programme for International Student Assessment (PISA).

### **Assessing adult populations**

Beginning in the 1990s, policy makers began to express a growing appreciation of the critical role that human capital, or the skills and knowledge that adults gain through education, workforce training, and lifelong learning, plays in outcomes for individuals and the societies in which they live. As a result, policy makers began to ask new sets of questions such as: What is the relationship between literacy skills and the ability to benefit from employer-supported training and lifelong learning? How are educational attainment and literacy skills related? How do literacy skills contribute to health and well-being as well as to participation and success in the labor force? What factors may contribute to the acquisition and decline of skills across age cohorts? How are literacy skills related to voting and other indices of social participation?

This growing policy interest led to a series of international assessments focusing on adults ages 16–65. The first of these large-scale, interview-based surveys was the International Adult Literacy Survey (IALS), which was conducted in multiple rounds from 1994 to 1999 in a total of 22 participating countries and sub-regions. This was followed by the survey of Adult Literacy and Lifeskills (ALL) in 11 countries and sub-regions.

Both IALS and ALL were designed to profile and explore the distribution of literacy skills among populations within and across participating countries. For each of these assessments, the construct of “literacy” was broadly conceived, being defined as: “Using printed and written information to function in society, to achieve one’s goals, and to develop one’s knowledge and potential” (Kirsch 2001). This definition, which traces its roots back to large-scale adult surveys conducted in the United States and Australia in

the late 1980s and early 1990s, characterized literacy as a set of complex information-processing skills that extend well beyond decoding and comprehending texts.<sup>2</sup> As a result, the assessment presented open-ended tasks for respondents to complete based on a range of intact, real-life stimulus materials in contexts ranging from health and safety, to personal finance, work, community resources, home and family, and consumer economics.

Tasks for IALS were developed around three domains, each representing a distinct and important aspect of literacy (Kirsch and Jungeblut 1986).

- *Prose literacy* the knowledge and skills needed to understand and use information from texts including editorials, news stories, poems, and the like;
- *Document literacy* the knowledge and skills required to locate and use information contained in job applications or payroll forms, bus schedules, maps, indexes, and so forth; and
- *Quantitative literacy* the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to complete tasks embedded in printed materials, such as balancing a check book, figuring out a tip, completing an order form, or determining the amount of interest on a loan from an advertisement.

In addition to these cognitive measures, IALS collected information about both the antecedents of skills in these domains as well as their outcomes through an extensive background questionnaire.

Like IALS, ALL included measures of prose and document literacy. However, the quantitative literacy domain from IALS was broadened to numeracy in this new assessment to reflect the evolving perspectives of experts in the field. Numeracy was defined as “the knowledge and skills required to effectively manage and respond to the mathematical demands of diverse situations” (Murray et al. 2005). The new numeracy domain was a more robust measure of a wider range of numerate behaviors, allowing ALL to collect more information about how adults apply mathematical knowledge and skills to real-life situations. In addition, ALL included a new domain focused on analytic problem solving skills—an area of growing policy interest given the importance of problem solving skills for success in the workplace and at home. Like IALS, ALL also included an extensive background questionnaire to collect data that allowed for an analysis of the relationships between skills and outcomes ranging from labor market participation and earnings, to physical and mental health, and engagement in community activities.

The work associated with developing and implementing IALS and ALL formed a knowledge base that contributed to the development and implementation of PIAAC in several important ways. Specific processes and procedures for the translation and adaptation of assessment instruments were developed and refined with the goal of ensuring comparability across language versions. Methods to evaluate the comparability of scoring within and across countries were established. Data analysis methodologies facilitated

<sup>2</sup> IALS adopted the definition of literacy used in the 1987 Young Adult Literacy Survey, the 1993 National Adult Literacy Survey in the United States, and an assessment of adult literacy conducted by the Commonwealth Department of Employment, Education and Training in Australia (Wickert 1989).

the evaluation of item-by-country interactions. And each assessment expanded what was measured, both in the context questionnaires and cognitive domains.

However, much like the new design for NAEP, which put large-scale assessment on a new trajectory, PIAAC marked a new and significant cycle of innovation. By moving to a computer-based assessment, PIAAC expanded what could be measured and improved the validity of large-scale adult assessment by including technology-based tasks that reflect the changing nature of information. In both the workplace and everyday life, it has become increasingly important that adults are able to navigate, critically analyze, and problem solve in data-intensive, complex digital environments—and PIAAC has made it possible to measure such skills. In addition, PIAAC has introduced methodological innovations such as multistage adaptive testing and flexible routing for the background questionnaire that have improved the design and delivery of the survey and laid the foundation for future assessments (Kirsch et al. 2017).

### The PIAAC assessment

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, incorporating 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.<sup>3</sup>

The first round of PIAAC included a Field Test designed to provide information related to four key areas.

- *Survey operations* including data collection procedures, response rates, and the efficiency and accuracy of data processing.
- *Instrument quality* focusing on the accuracy and comparability of survey instruments including translation and scoring guides, the timing and flow of questions in the background questionnaire, and the appropriateness of questions across participating countries.
- *Platform* focusing on the computer platform in terms of response capturing and automatic scoring, functioning of the computer-assisted personal interviewing (CAPI) system, accuracy of instructions for the interviewer, and the integration of the PIAAC platform with national survey management systems.
- Psychometric characteristics of the items and scales, including the equivalence of item parameters between paper-and-pencil and computer formats.

The Field Test was also used to examine the role of computer familiarity and to determine the standards for routing respondents to the paper instruments. Data from the Field Test provided the initial IRT parameters used to construct the adaptive testing

<sup>3</sup> Reading 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. And most implemented problem solving, with the exceptions being France, Italy and Spain. In Rounds 2 and 3, there were no optional components and these two domains were treated as core components.



algorithms that were then implemented in the Main Study. The outcomes of the Field Test were used to assemble the final instruments and modify or refine any operational issues in order improve the overall quality of the Main Study.

### **Background questionnaire**

As was the case in IALS and ALL, 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 these surveys, which has been to relate skills to a variety of demographic characteristics and explanatory variables. The information collected via the background questionnaire adds to the interpretability of the assessment, thereby enhancing the reporting of results to policy makers 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 an understanding of the factors related to the observed distribution of literacy skills across populations as well as factors that mediate the acquisition or decline of those skills.

Background information also contributes to the psychometric modeling of the data by providing auxiliary information that can be used to improve the precision of the skills measurement. This use of background data is particularly important because it permits the use of assessment designs in which each respondent need only receive a subset of the full item pool developed for each domain while also optimizing the estimation of proficiency for a population or subpopulation of interest.<sup>4</sup>

A major benefit of using a computerized questionnaire in PIAAC is the application of flexible routing so that parts of the questionnaire can be skipped in order to tailor a question, or block of questions, to an individual or group of respondents. Based on response patterns, variables can be derived that control the flow of the questionnaire. For example, only those respondents who reported that they had been looking for work were asked about the methods they were pursuing to find employment; those who reported they were not looking for work received additional questions about reasons for not doing so. This modular approach allowed more flexibility in the use of the allotted assessment time and helped reduce respondent burden.

### **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 taken from IALS and ALL. In order to maintain trend measurement, the PIAAC design required that 60% of the literacy and numeracy items be taken from previous surveys, with the remaining 40% being newly developed items. 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.

---

<sup>4</sup> The interested reader is referred Mislevy et al. (1992) for a description of this approach and to von Davier et al. (2006) for an overview and a description of recent improvements and extensions of the approach.



To establish common scales for the literacy and numeracy items, those items had to be linked across assessment modes for PIAAC. This was achieved by using common sets of items in both modes in the Field Test. Respondents were administered a brief screener that assessed their ability to click, type a single-word response, select from a drop-down menu, scroll, drag and drop, and highlight. Those who passed were randomly assigned to either the paper or computer instruments, a design that made it possible to evaluate the extent to which item parameters were consistent across modes for each domain. The Field Test scaling analysis revealed that there was overwhelming consistency across modes for both the literacy and numeracy linking items so that a single common scale could be established for each domain that was linked across both time and mode of assessment.<sup>5</sup>

The primary considerations when selecting linking items for PIAAC included item quality, fit with the framework dimensions, distribution across levels of difficulty, and cultural appropriateness for participating countries. Additionally, trend materials needed to be evaluated in terms of suitability for computer delivery as they had all been originally designed for paper-and-pencil administration. Stimulus materials needed to be adaptable to an onscreen presentation, keeping the same formatting as that used on paper, and all selected items needed remain open-ended, but be capable of being computer scored in order to support the adaptive design of PIAAC. The development consortium relied on evidence from previous ETS work on a derivative computer-based test for individuals to define a set of computer-scoreable, open-ended response modes for the trend items. This work had shown that item parameters for paper-and-pencil items were not impacted when those items were adapted to allow respondents to click on answers, type numeric responses, and highlight answers in a text. Development therefore proceeded on the assumption that linking items could be adapted to employ these response modes and still maintain item parameters from previous assessments, an assumption that was ultimately supported by the Field Test data (OECD 2013).

The four cognitive domains are explained in more detail below. Literacy and numeracy items were included in both the paper- and computer-based versions of the assessment, reading components was paper-based only, and problem solving in technology-rich environments was developed solely as part of the computer-based instrument.

- Literacy
- The PIAAC literacy scale included both prose and document literacy tasks.<sup>6</sup> While literacy had been a focus of both the IALS and ALL surveys, PIAAC was the first of these surveys to address literacy in digital environments. As a computer-based assessment, PIAAC included literacy tasks that required respondents to use electronic texts including web pages, e-mails, and discussion boards. These interactive stimulus materials included hypertext and multiple screens of information and simulated real-life literacy demands presented by digital media.

<sup>5</sup> See Chapters 17 and 18 in the PIAAC Technical Report (OECD 2013) for a more detailed explanation of how the scales were linked across delivery modes and surveys.

<sup>6</sup> While the IALS and ALL surveys included separate prose and document literacy scales, those domains were rescaled to form a single literacy scale for PIAAC.

- 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 behavior for PIAAC: *Numerate behavior 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 behavior was further specified as follows.
  1. *Real contexts* include everyday life, work, society, and further learning.
  2. *Responding* may require any of the following: identifying, locating or accessing, acting upon and using (to order, count, estimate, compute, measure or model), interpreting, evaluating or analyzing, and communicating mathematical content, information or ideas.
  3. *Mathematical content, information, and ideas* include: quantity and number, dimension and shape, pattern, relationships and change, and data and chance.
  4. *Representations* may include: objects and pictures, numbers and mathematical symbols, formulae, diagrams, maps, graphs and table, texts, and technology-based displays.
- 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, and fluency in the reading of short passages of text.
- 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). Assessing reading component skills was important in the evolution of adult surveys because in order to have a full picture of literacy in any society it is necessary to have more information about those individuals who are at the greatest risk of negative social, economic, and labor market outcomes.
- Problem solving in technology-rich environments (PSTRE)
- PSTRE was a new domain introduced in PIAAC and represented the first attempt to assess this domain 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 solving information problems within the context of information and communication technologies (ICT) rather than on analytic problems per se. PSTRE 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 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).

- The PSTRE 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.

#### ***How skills were measured***

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 web pages 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 that these items could be automatically scored for the first time, which contributed to improved scoring reliability within and across countries. Three open-ended item formats were included:

- Clicking items
  - Respondents were asked to click on graphical elements, cells in a table, links on a web page, or radio buttons or check boxes to answer.
- Numeric entry items
  - Respondents answered by typing a numeric response using the number keys, decimal point (represented using a period or comma as appropriate for each participating country) and space key. In this response mode, all other keys on the keyboard were locked to prevent respondents from including text in their responses that could not be automatically scored. Numeric entry items could be scored automatically based on the definition of correct numeric responses included in the scoring rules.
- Highlighting items
  - Respondents were able to freely highlight one or more words, phrases and sentences in a text to answer questions. Developers defined a minimum correct response, as well as a maximum correct response, for each highlighting item. These judgments were based on ETS’s previous development of open-ended, computer-scoreable items as well as experience with paper-based versions of these items, where scoring rules had been developed to take into consideration instances where respondents underlined or circled information in the stimulus instead of writing an answer on the provided response line.

In addition to being computer scoreable, each of these three formats required only basic computer skills—an important consideration given that the test needed to be accessible to adults with varying degrees of computer experience.

### Assessment design

PIAAC is a household survey, meaning that it is administered in face-to-face interviews in the homes of nationally representative samples of adults. It was designed as a computer-based survey, with interviewers bringing laptops into participants' homes. While the primary mode of administration was computer, a paper mode was developed as well. In the Main Study, adults who were either unable or unwilling to use a computer were provided with paper-and-pencil assessment booklets.

As can be seen in Fig. 2, the mode of administration was determined by responses to questions about ICT use in the background questionnaire (BQ), performance on an ICT screener, and performance on a cognitive screener.

Those respondents who reported some computer familiarity, passed the two screeners, and were willing to do so, took the assessment on the computer. In nearly all countries across Rounds 1 and 2 of PIAAC, the majority of respondents were in this category.<sup>7</sup> Those respondents who took the paper version comprised three groups:

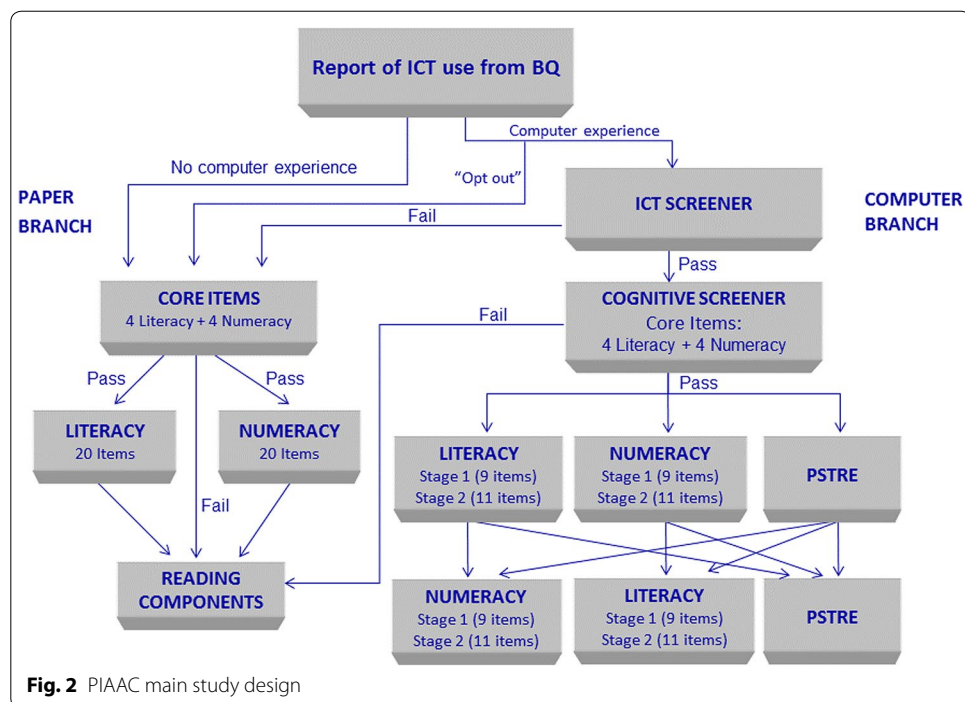
- a. adults who reported in the BQ that they did not use a computer at home or work (e.g., they did not use email, the internet, make purchases, bank, use spreadsheets, use a word processor, write programs or use social media);
- b. adults who reported that they had computer experience but “opted out,” or refused to take the computer-based version of the assessment; and
- c. those who reported that they had computer experience but were unable to demonstrate basic computer skills as assessed via the ICT screener where they were asked to click, type a single-word response, select from a drop-down menu, scroll, drag and drop, and highlight.

As shown in Fig. 2, participants in these three groups were administered paper booklets with literacy or numeracy tasks followed by the assessment of reading component skills.<sup>8</sup> Any participants who failed the paper-based core tasks (consisting of 4 literacy and 4 numeracy items) were routed to the reading components assessment. One additional group, made up of respondents who passed the ICT screener but failed the cognitive screener, were given just the paper-based measure of reading components.<sup>9</sup>

<sup>7</sup> Reports of participation in the computer-based assessment and the paper-based assessment by country in Round 1 can be found in Section A7-3 (adjudication reports, assessment data section) of the PIAAC Technical Report, (OECD 2013).

<sup>8</sup> On average across all countries in Rounds 1 and 2 some 9–10% of respondents were in each of groups (a) and (b). Less than 5% of respondents were in group (c). See the Reader's Companion for The Survey of Adult Skills (OECD 2016) for more detailed information.

<sup>9</sup> This group included less than 1% of participants in the survey. See the Reader's Companion for The Survey of Adult Skills (OECD 2016) for more detailed information.

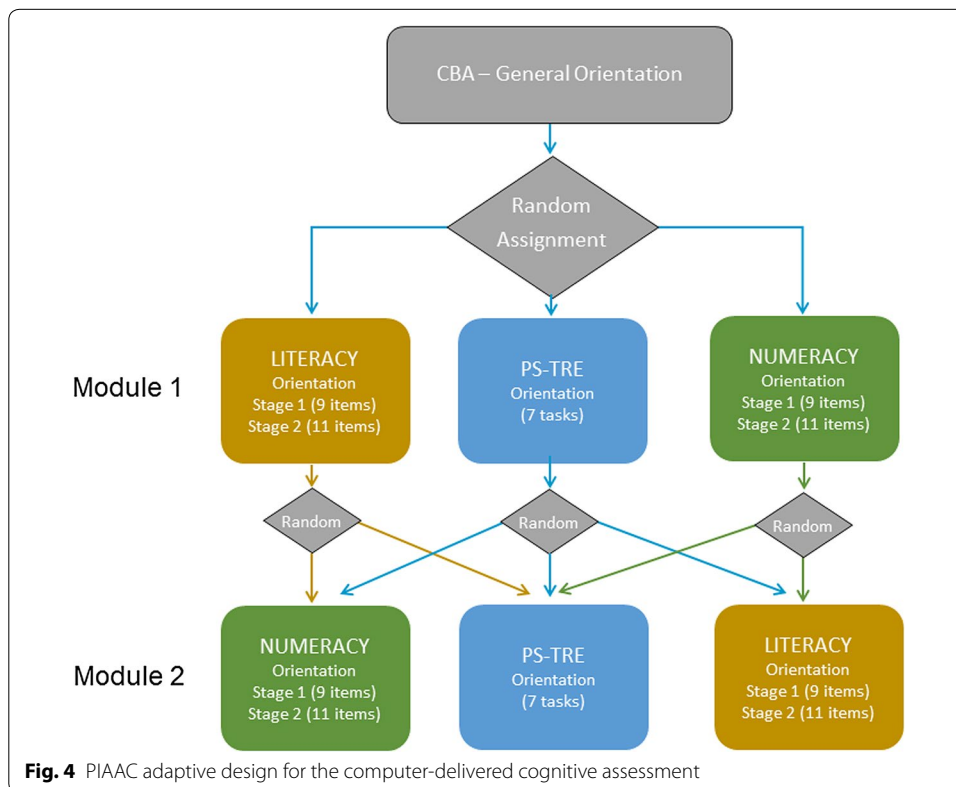
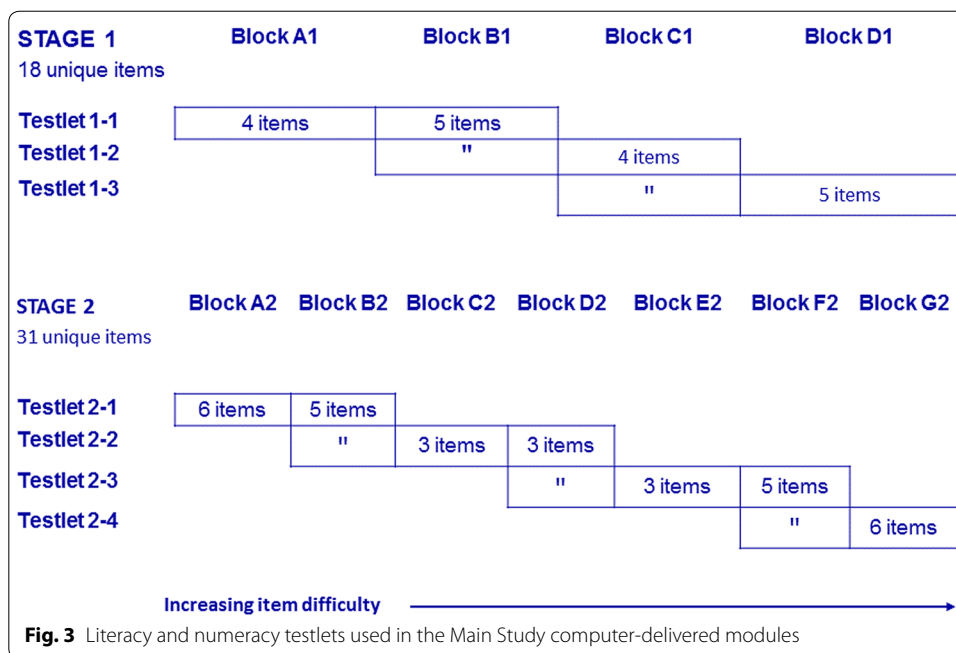


### Multistage adaptive testing

The computer-based assessment environment used in PIAAC made it possible to implement an assessment design that included multistage adaptive testing. This is a variant of item-level adaptive testing, in which a response to a single item determines the next item presented. The multistage design algorithms work on a testlet, or cluster, level where responses to a number of items determine the next testlet presented to the test taker. This design makes it possible to collect more performance information and therefore increases the selection accuracy for the next testlet. As noted previously, data from the Field Test provided the initial IRT parameters that were used to construct the adaptive testing algorithm that was then implemented in the Main Study.

The literacy and numeracy domains in the cognitive assessment were designed around two stages with a total of seven testlets: three in stage 1 and four in stage 2, as shown in Fig. 3. The set of items presented to a given respondent in Stage 1 was based on background variables collected in the BQ, as well as the score received on the cognitive screener. Stage 1 included only four blocks of items that broadly covered the range of item difficulty as this initial routing decision was based on limited information. The testlet assigned in Stage 2 was based on background variables, the cognitive screener and the respondent's performance on the set of items administered in Stage 1. The increased amount of available information made more precise assignments possible in Stage 2, where each block of items covered a narrower range of the difficulty spectrum. More able respondents received a more difficult set of items than less able respondents. This design optimized the match between item difficulty and respondent ability, providing more reliable information about a respondent's skills within the specified testing time.

The overall design for the Main Study computer-based assessment is shown in Fig. 4. In Module 1 of the cognitive assessment, respondents were randomly assigned to either



the literacy, numeracy or PSTRE domain. Those assigned to literacy or numeracy took both stages of those assessments, receiving a total of 20 items. In Module 2, those respondents who received literacy in Module 1 were randomly assigned to either

numeracy or PSTRE. Those who started with numeracy in Module 1 were randomly assigned to either literacy or PSTRE; and those who took PSTRE were randomly assigned to either literacy, numeracy or a second module of PSTRE.<sup>10</sup>

### Scaling and comparing proficiencies

Across the computer-based and paper-based instruments, a total of 58 literacy, 56 numeracy and 14 PSTRE tasks were administered to nationally representative samples of adults in each participating country to ensure the broadest possible coverage of each domain given the constraints of the study. Because no single adult could be expected to respond to the entire set of tasks, the design for PIAAC required that each participant receive and respond to a subset of tasks from each of the three cognitive domains.

Summarizing the performance of adults across the entire set of tasks posed a challenge. To establish a common scale for each of the domains, tasks first had to be carefully assembled into testlets that linked across modes and across surveys.<sup>11</sup> This was accomplished following the assessment design presented earlier in this paper ensuring that each set of tasks was administered to representative samples in each country. Once the data were collected, the pool of tasks within each domain was analyzed in a way that would array the set of tasks along a continuum that both reflected the proficiency of adults in a particular domain as well as the level of skill and knowledge associated with a correct response. As discussed earlier, the procedure used in PIAAC was IRT- based.

PIAAC used the two-parameter logistic model (2PL; Birnbaum 1968) for dichotomously scored responses and the generalized partial credit model (GPCM; Muraki 1992) for items with more than two response categories. The 2PL model is a mathematical model for the probability that an individual will respond correctly to a particular item from a single domain of items. The probability of solving an item depends only on the respondent's ability, or proficiency, and two item parameters characterizing the properties of the item (item difficulty and item discrimination). This model was used to calibrate the items for each domain as well as to link items across modes and across surveys.

Once a fixed set of international and national item parameters was established, a latent regression model was fitted to the data and plausible values were estimated for each respondent in each country. Plausible values are multiple imputed proficiency values based on information from the test items (the actual PIAAC literacy, numeracy, and PSTRE instruments) and information provided by the respondent in the BQ. Plausible values are used to obtain more accurate estimates of group proficiency than would be obtained through an aggregation of point estimates. More detailed information describing the procedures used to scale the cognitive data and estimate proficiency values for each respondent can be found in the technical report available electronically through the OECD (OECD 2013).

<sup>10</sup> See Chapter 1, PIAAC Assessment Design, in the PIAAC Technical Report (OECD 2013) for a more detailed explanation of the adaptive routing procedures.

<sup>11</sup> As shown in Fig. 3, the computer-based items were organized into testlets. In the paper-based instruments, items were assembled into clusters, as described in Annex A1 of Technical Report (OECD 2013).



### Creating described proficiency scales and reporting results

Although creating the three scales used to assess proficiency in PIAAC was a major goal of the survey, the numerical scores themselves carry little or no meaning. For example, while most people have a practical understanding of the weather and how they should dress when the temperature is at 10 °C, it is not obvious what it means when a particular group or subgroup in a country is shown to score at 254 on the numeracy scale or 263 on the literacy scale.

One way to develop an understanding about what a particular score along a scale means is to compare one group within a country to another—such as comparing the average score of people who are employed full time with those who are unemployed, or the average score of those who completed secondary education with those who did not. Clearly, comparing groups within and across countries on selected variables is one meaningful way to gain some understanding of how performance is distributed and connected to outcomes of interest, but this approach doesn't help explain what is being assessed from a construct point of view. A deeper understanding requires focusing on the underlying construct and how it has been measured in a particular survey.

PIAAC, like most large-scale surveys, relies on one or more groups of experts to guide the development of instruments. This guidance is provided through the development of a framework for each of the domains. The overall purpose of a framework is to enhance measurement by identifying key features of each domain that must be reflected in the item pool.

Experts for each of the three PIAAC cognitive domains employed a consensus-building process to develop and adopt a working definition for literacy, numeracy, and PSTRE. In operationalizing these definitions, they specified key task characteristics associated with critical features necessary to demonstrate proficiency in that domain. For example, the model for PIAAC literacy included text features, aspects of tasks, and a range of content areas or social contexts from which the texts were to be selected. Once identified, these task characteristics were specifically defined and used by test developers to create items that could be mapped back to the framework. At the completion of the test development process, the experts met again to review the items, confirm their framework classifications, and approve items for the Field Test. The expert groups met for a final time after the Main Study to review the results in order to create or refine descriptions of proficiency along each of the scales. These descriptions relied both on the task characteristics that were used to guide item development and the location of items along the continuum of each scale that was based on the item calibration process and the selection of a response probability, or RP, value.

Along with the task characteristics, the RP value chosen to characterize items along each scale helps to define what is meant by proficiency in PIAAC. It was decided that “proficiency” for the purposes of PIAAC should mean that respondents would have a 67% chance of correctly answering all items at the same point on the scale. This means that any adult with an estimated proficiency of 275 would have a 67% chance of responding correctly to all items at 275 on that scale. This should not be taken to mean that adults who scored below 275 would always respond incorrectly or that adults slightly above this point would always get the item correct. Rather, adults at different points on the scale have a greater or lesser chance of responding to an item at 275 correctly or

incorrectly. It also means that adults would have a higher chance of responding correctly to all tasks that are easier, or below 275 on the scale, and a lower chance of responding correctly to items that are above 275 on the scale.

More information about the frameworks for each of the cognitive scales including the task characteristics and described proficiency levels that were developed or refined in conjunction with each of the expert groups are available electronically at the OECD website.<sup>12</sup>

## Data products

A complex survey such as PIAAC generates an extensive volume of data of interest to a wide range of users. To support the dissemination and analysis of the PIAAC data, a number of data products have been developed, including the following.

- *Data Explorer*<sup>13</sup> a web-based analysis and reporting tool that permits users to query the PIAAC database and produce presentation-quality tabular and graphical summaries of the data. This tool has been designed for a wide range of potential users, including those with little or no statistical background. Both private versions, for use by the OECD and participating countries, and public versions of the Data Explorer are provided. The Data Explorer includes all released international and national variables.
- *Summary tables* a comprehensive set of tables that contain weighted summary statistics for each participating country on each cognitive item and each variable in the background questionnaire. The public version of the summary tables is the “Data Compendia”.<sup>14</sup> As described on the OECD web site: “The compendia are sets of tables that provide categorical percentages for both cognitive and background items. The purpose of the compendia is to support users of the public use file (PUF) so that they can gain knowledge of the contents of the PUF and can use the compendia results to be sure that they are performing PUF analyses correctly. Note that due to the design of the cognitive assessment, comparisons of the cognitive item statistics provided in the compendia across countries for reporting purposes may not be appropriate.”
- *Public use data*<sup>15</sup> a web-based delivery system for data files and client-based data management and analysis tools that a wide range of users can operate on their own computer systems. To protect the confidentiality of individuals, any personally-identifiable information is excluded from the public use data products. In addition, the system complies with all national reporting regulations such as those that require that only suppressed or coarsened data be included.
- *Electronic codebook*<sup>16</sup> a client-based Windows application for use with either the international database or the public use data to supplement the variable selection and data analysis functions of the Data Explorer. The program allows the end user to

<sup>12</sup> For information about the frameworks, see OECD (2012). A more detailed description of the PIAAC proficiency scales can be found in Chapters 19 and 21 in the PIAAC Technical Report (OECD 2013).

<sup>13</sup> The Data Explorer can be accessed at: <http://piaacdataexplorer.oecd.org/ide/idepiac/>.

<sup>14</sup> The Data Compendia can be accessed at: <http://www.oecd.org/skills/piaac/publicdataandanalysis/>.

<sup>15</sup> The public use data products can be accessed at: <http://www.oecd.org/skills/piaac/publicdataandanalysis/>.

<sup>16</sup> The electronic codebook can be accessed at: <http://www.oecd.org/skills/piaac/publicdataandanalysis/>.

view the attributes of the variables in a data set of interest and select a subset of variables for use in analysis. Optional outputs of the program include an extract data file consisting of only the variables and cases of interest and syntax files for creating data files for a number of popular data analysis systems including, but not limited to, SPSS, SAS, STATA, and R. The application can also be bundled with a library of macros for each of those systems to perform appropriate analyses of the data within those systems.

- *International Database (IDB) Analyzer*<sup>17</sup> this application, developed by the IEA's Data Processing and Research Center, facilitates the analysis of large-scale assessment data. The tool allows users to conduct statistical analyses taking into account the complex sampling design structure of the PIAAC database, which cannot be handled correctly by SPSS alone. The IDB Analyser generates SPSS syntax that fully takes into account information from the participant's sampling design in the computation of sampling variance. In addition, it handles plausible values. The software allows users to combine data from different countries for cross-country analysis and to select specific subsets of variables.

In addition to these specific data products, additional materials that are made publicly available include the technical report, assessment frameworks, and sample items. The technical report (OECD 2013) is written by members of the consortium responsible for developing and delivering the assessment and scaling and analyzing the results. This document provides readers with information about the assessment design, instrument design and development, translation, platform development, field operations, sampling and weighting, and data analysis and results.

The frameworks (OECD 2012) are written by the expert groups for each domain and provide an overview of that domain, define the construct, the performances or behaviors expected to reveal that construct, and the characteristics of the assessment tasks to elicit those behaviors. The framework provides a detailed blueprint about what is to be measured and how results will be interpreted and reported. By explicitly describing the focus of the assessment, the framework documents provide valuable information about what the assessment is, and is not, intended to measure and thus how PIAAC is similar to, and different from, other assessments. Similarly, the sample items provide examples of how the frameworks were instantiated through the test items and help provide a clearer picture of the assessment.

### Extending the utility of PIAAC

The development and conduct of a large-scale assessment such as PIAAC is an enormous undertaking involving literally thousands of individuals—from survey participants to interviewers, staff at the national centers and survey organizations in participating countries, and members of the consortia responsible for the design, development and conduct of the survey—all under the direction of the Board of Participating Countries and OECD. Ensuring that the data are sound and then that they are widely available and accessible to interested parties is the critical final stage in the life cycle of such an effort.

<sup>17</sup> The IDB Analyzer can be accessed at: <http://www.oecd.org/skills/piaac/publicdataandanalysis/>.

The data products and analysis tools described above have provided unprecedented access to the expansive PIAAC database. To promote the appropriate use and analysis of the data, some 15 workshops have been conducted internationally over the past three years to provide training around the data products and the structure of the PIAAC data. Interest in these data is widespread and ongoing, as evidenced by published analyses of the data and the development of derivative products.

### Secondary level policy analyses

To date, some 200 reports focusing on the PIAAC data have been published. They address a wide range of topics, including: skill patterns, differences in skills among subgroups such as youth and immigrants, returns to skills in the labor market, lifelong learning, ICT skills, wage and income inequality within and across countries, adult education and training, skills and social outcomes such as health, trust and cultural participation, and policy interventions and implications of the PIAAC findings. The range of disciplines utilizing the PIAAC databases for secondary analyses is broad and reflected in the journals and magazines publishing this work, some of which include: *The Journal of Education Finance*, *Educational Studies*, *Journal of Social Policy Studies*, *European Educational Research Journal*, *Advances in Social Sciences Research Journal*, *The Economist*, *Computers & Education*, *The Journal of Policy Modeling*, *Sociology of Education*, and *International Review of Education*.

Additionally, national reports focusing on country-level data and comparisons across countries, as well as methodological reports, have been developed. Many have been published by the OECD as well as by national bureaus and research organizations.<sup>18</sup> ETS, as but one example, established a policy center to conduct secondary level analyses using data from large-scale assessments. Its work associated with PIAAC has included additional analyses of the reading component skills of adults in the U.S. (Sabatini 2015) as well as an analysis of the skills of America's millennials as they compare with those of their international peers (Goodman et al. 2015).

To further support analysis and dissemination efforts, three international conferences have been held to promote the use of PIAAC data for addressing policy issues. Taken together, these workshops, publications and conferences reflect the importance of the PIAAC data across a range of disciplines including education, labor economics, sociology and social policy.

### Derivative products: Education & Skills Online

Finally, the work of large-scale assessments can be further extended through derivative products that make use of the content, development processes and procedures, and data from the assessment for new purposes. For example, national large-scale assessments in the U.S. in the 1990s formed the basis for several derivative products including: the Test of Applied Literacy Skills (TALS), a paper-and-pencil test that yielded individual-level results; a multi-media group-based instructional system for adults that focused on

<sup>18</sup> See a full list of PIAAC international reports and working papers published by the OECD at <http://www.oecd.org/skills/piaac/publications.htm>.

prose, document and quantitative literacy; and the PDQ Profile series, an adaptive computer-based assessment of literacy proficiency for individuals.

Following that same model, Education & Skills Online (ESOL) was developed as an online adaptive assessment designed to provide individual-level results that are linked to PIAAC. Measures of literacy and numeracy are included in this derivative product, as well as optional assessments of reading components and problem solving in technology-rich environments. Because of its link to PIAAC, results from ESOL can be benchmarked against national and international results for participating countries. An optional assessment of non-cognitive skills is also included in the product.

The primary purpose of ESOL is to provide information about the skills of individuals, either to inform training efforts or for research purposes. As such, the OECD identifies potential users as follows (“Education & Skills Online Assessment,” n.d.):

- “Organisations providing adult literacy and numeracy training that wish to have information that can help diagnose the strengths and weaknesses of learners and evaluate the results of training against national and international benchmarks.
- Educational institutions such as universities, vocational education and training centers that can use *Education & Skills Online* as a diagnostic tool for incoming students to help determine their need for literacy/numeracy courses.
- Researchers who would like to have access to an assessment that is benchmarked to PIAAC results.
- Government organisations interested in assessing the learning needs of unemployed adults, at risk groups or economically disadvantaged adults.
- Public or private companies that want to use the results to help them identify the training needs related to literacy and numeracy for their workforce.”

## Summary

Like the new design for NAEP in the era of the 1980s, the Programme for the International Assessment of Adult Competencies marked a turning point in the almost 25-year history of international large-scale assessments of adults. In many ways, PIAAC represented the culmination of all that was learned over the several preceding decades in terms of instrument design, translation and adaptation procedures, scoring of open-ended items, and the development of interpretive schemes for large-scale assessments. But in response to new policy questions, and as the first computer-based survey of adult skills, PIAAC also made it possible to introduce significant innovations. These included:

- Multistage adaptive testing;
- Automated routing for the background questionnaire and a complex design for the cognitive assessment;
- Fully automated scoring of open-ended items across more than 50 language versions of the assessment;
- Expansion of what could be measured in the existing constructs—for example, by including electronic texts and interactive stimulus materials;
- Addition of new constructs including reading components, which added better measurement at the lower end of the literacy scale, and problem solving in technol-

ogy-rich environments, which challenged respondents to solve open-ended information problems in ICT environments;

- Inclusion of new item types and response modes; and
- Use of extensive log files to improve data interpretation.

Such innovations reflect a new era of increasing literacy demands as the types and amount of information adults must manage in their daily lives continue to expand.

The impact of PIAAC has grown as policy makers and other stakeholders increasingly come to appreciate the critical role that skills play in allowing individuals to maintain and enhance their ability to meet changing work conditions and societal demands. The PIAAC data provide a better understanding of the distribution of those key skills and proficiencies at both national and international levels. They shed light on the extent to which skills translate into better opportunities and outcomes for individuals and into stronger economies. And they inform the evaluation of the effectiveness of our education and training systems, as well as our social and workplace practices, in developing required skills and proficiencies.

As the largest and most innovative survey of adult skills ever conducted, PIAAC both complemented and broadened the types of information collected in school-based surveys. The innovation introduced via PIAAC increased the relevance of the survey along with the accuracy of the data. As such, PIAAC contributed to improved relevance, quality and validity in large-scale assessments.

#### Abbreviations

ALL: Adult Literacy and Lifeskills; BIB: balanced incomplete block; BQ: background questionnaire; CAPI: computer-assisted personal interviewing; ESOL: education and skills online; ETS: Educational Testing Service; IALS: International Adult Literacy Survey; ICT: information and communication technologies; IDB: international database; IEA: International Association for the Evaluation of Educational Achievement; IRT: item response theory; NAEP: National Assessment of Educational Progress; OECD: Organisation for Economic Co-operation and Development; PIAAC: Programme for the International Assessment of Adult Competencies; PIRLS: Progress in International Reading Literacy Study; PISA: Programme for International Student Assessment; PSTRE: problem solving in technology rich environments; PUF: public use file; RP: response probability; TALS: Test of Applied Literacy Skills; TIMSS: Trends in International Mathematics and Science Study; 2PL: two-parameter logistic model.

#### Authors' contributions

The authors co-authored the work. Both authors read and approved the final manuscript.

#### Acknowledgements

Not applicable

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 28 March 2017 Accepted: 4 April 2017

Published online: 18 April 2017

#### References

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.
- Education & Skills Online Assessment (n.d.). <http://www.oecd.org/skills/ESonline-assessment/abouteducationskillsonline>. Accessed 23 Feb 2017.

- Goodman, M., Sands, A., & Coley, R. (2015). *America's skills challenge: Millennials and the future*. Retrieved from Educational Testing Service Research Website: <https://www.ets.org/s/research/30079/asc-millennials-and-the-future.pdf>. Accessed 15 Feb 2017.
- Grenier, S., Jones, S., Strucker, J., Murray, T. S., Gervais, G., & Brink, S. (2008). *Learning literacy in Canada: Evidence from the International Survey of Reading Skills*. Ottawa: Statistics Canada.
- Kirsch, I. (2001). *The International Adult Literacy Survey (IALS): Understanding what was measured* (Research Report No. RR-01-25). Princeton, NJ: Educational Testing Service.
- Kirsch, I. S., & Jungeblut, A. (1986). *Literacy: profiles of America's young adults* (NAEP Report No. 16-PL-01). Princeton, NJ: Educational Testing Service.
- Kirsch, I., Lennon, M., von Davier, M., Gonzalez, E., & Yamamoto, K. (2013). On the growing importance of international large-scale assessments. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The role of international large-scale assessments: Perspectives from technology, economy, and educational research*. New York: Springer.
- Kirsch, I., Lennon, M., Yamamoto, K. & von Davier, M. (2017, in press). Large-scale assessments of adult literacy. In R. Bennett & M. von Davier (Eds.), *Advancing human assessment: Methodological, psychological, and policy contributions*. New York: Springer.
- Messick, S., Beaton, A., & Lord, F. (1983). *National Assessment of Educational Progress reconsidered: A new design for a new era* (NAEP Report 83-01). Princeton, NJ: Educational Testing Service.
- Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133–161.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159–177.
- Murray, T. S., Clermont, Y., & Binkley, M. (Eds.). (2005). *Measuring adult literacy and life skills: New frameworks for assessment* (Report 89-552-MIE, No. 13). Ottawa: Statistics Canada.
- Naemi, B., Gonzalez, E., Bertling, J., Betancourt, A., Burrus, J., Kyllonen, P., et al. (2013). Large-scale group score assessments: Past, present, and future. In D. Saklofske, V. Schwean, & C. R. Reynolds (Eds.), *Oxford handbook of child psychological assessment* (pp. 129–149). Cambridge, MA: Oxford University Press.
- OECD (2012). *Literacy, numeracy and problem solving in technology-rich environments: Framework for the OECD Survey of Adult Skills*. Retrieved from OECDiLibrary Website: <http://dx.doi.org/10.1787/9789264128859-en>. Accessed 15 Feb 2017.
- OECD. (2013). *Technical report of the Survey of Adult Skills (PIAAC)*. Retrieved from OECDiLibrary Website: [https://www.oecd.org/skills/piaac/\\_Technical%20Report\\_17OCT13.pdf](https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf). Accessed 15 Feb 2017.
- OECD. (2016). *The Survey of Adult Skills: Reader's companion* (2nd ed.). Retrieved from OECDiLibrary Website: [http://www.oecd-ilibrary.org/education/the-survey-of-adult-skills\\_9789264258075-en](http://www.oecd-ilibrary.org/education/the-survey-of-adult-skills_9789264258075-en). Accessed 23 Feb 2017.
- Sabatini, J. (2015). *Understanding the basic reading skills of U.S. adults: Reading components in the PIAAC literacy survey*. Retrieved from Educational Testing Service Research Website: <https://www.ets.org/s/research/report/reading-skills/ets-adult-reading-skills-2015.pdf>. Accessed 15 Feb 2017.
- 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 (Report No. 29). Boston, MA: National Center for the Study of Adult Learning and Literacy.
- von Davier, M., Sinharay, S., Oranje, A., & Beaton, A. (2006). The statistical procedures used in the National Assessment of Educational Progress: Recent developments and future directions. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26, pp. 1039–1055)., Psychometrics Amsterdam: Elsevier.
- Wickert, R. (1989). *No single measure: A survey of Australian adult literacy*. Canberra: The Commonwealth Department of Employment, Education and Training.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)