Chapter 10
Analysing Log File Data from PIAAC

Frank Goldhammer, Carolin Hahnel, and Ulf Kroehne

Abstract The OECD Programme for the International Assessment of Adult Competencies (PIAAC) was the first computer-based large-scale assessment to provide anonymised log file data from the cognitive assessment together with extensive online documentation and a data analysis support tool. The goal of the chapter is to familiarise researchers with how to access, understand, and analyse PIAAC log file data for their research purposes. After providing some conceptual background on the multiple uses of log file data and how to infer states of information processing from log file data, previous research using PIAAC log file data is reviewed. Then, the accessibility, structure, and documentation of the PIAAC log file data are described in detail, as well as how to use the PIAAC LogDataAnalyzer to extract predefined process indicators and how to create new process indicators based on the raw log data export.

The Programme for the International Assessment of Adult Competencies (PIAAC) is an Organisation for Economic Co-operation and Development (OECD) study that assesses and analyses adult skills in the cognitive domains of literacy, numeracy, and problem solving in technology-rich environments (PS-TRE). The computer-based assessment requires respondents to solve a series of tasks (items). The tasks are related to information presented to the respondent on the screen (e.g. a text from a newspaper, a simulated webpage). When solving a task, the respondent interacts with the assessment system—for example, by entering or highlighting text or clicking graphical elements, buttons, or links. The assessment system logs all these interactions and stores related events (e.g. keypress) and time stamps in log files.

F. Goldhammer (✉) · C. Hahnel · U. Kroehne

DIPF | Leibniz Institute for Research and Information in Education, Centre for International Student Assessment [ZIB], Frankfurt, Germany

e-mail: goldhammer@dipf.de

For the cognitive assessment in Round 1 of PIAAC (2011–2012), the OECD has provided both the log data and a supporting infrastructure (i.e. the extraction tool PIAAC LogDataAnalyzer and online documentations) to make the log data accessible and interpretable (OECD 2019). Overall, 17 of the participating countries (i.e. Austria, Belgium [Flanders], Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Korea, the Netherlands, Norway, Poland, Slovak Republic, Spain, the United Kingdom [England and Northern Ireland], and the United States) agreed to share their log data with the research community. Log file data are available from the computer-based assessment of all three PIAAC cognitive domains—literacy, numeracy, and PS-TRE—but not for the background questionnaire. The log file data extend the PIAAC Public Use File, which mainly includes the result data (i.e. scored item responses) of the individual respondents from the cognitive assessment and the background questionnaire.

The goal of this chapter is to familiarise researchers with how to access, understand, and analyse PIAAC log file data for their research purposes. Therefore, it deals with the following conceptual and practice-oriented topics: In the first part, we will provide some conceptual background on the multiple uses of log file data and how to infer states of information processing by means of log file data. The second part reviews existing research using PIAAC log file data and process indicators included in the PIAAC Public Use File. The third part presents the PIAAC log file data by describing their accessibility, structure, and documentation. The final part addresses the preprocessing, extraction, and analysis of PIAAC log file data using the PIAAC LogDataAnalyzer.

10.1 Log File Data Analysis

10.1.1 Conceptual Remarks: What Can Log File Data from Technology-Based Assessments Be Used For?

The reasons for using log file data in educational assessment can be diverse and driven by substantive research questions and technical measurement issues. To classify the potential uses of log file data from technology-based assessments, we use models of the evidence-centred design (ECD) framework (Mislevy et al. 2003). The original ECD terminology refers to the assessment of students, but the approach is equally applicable to assessments of the general adult population (such as PIAAC). The ECD framework is a flexible approach for designing, producing, and delivering educational assessments in which the assessment cycle is divided into models. The ECD models of interest are (a) the student model, (b) the evidence model, (c) the task model, and (d) the assembly model. Applying the principles of ECD means specifying first what construct should be measured and what claims about the respondent are to be made based on the test score (student model). Then, the type of evidence needed to infer the targeted construct and the way it

can be synthesised to a test score across multiple items is explicated (evidence model). Based on that, the items are designed in such a way that they can elicit the empirical evidence needed to measure the construct (task model). Finally, the items are assembled to obtain a measurement that is reliable and that validly represents the construct (assembly model). Although the ECD framework was originally developed for assessments focusing on result or product data (i.e. item scores derived from the respondent's work product), it is suitable for identifying the potential uses of log file data in the fields of educational and psychological assessment, as shown in the following.

The *student model,* (a), addresses the question of what latent constructs we want to measure (knowledge, skills, and attributes) in order to answer, for example, a substantive research question. Thus, one reason to use log file data is to measure constructs representing attributes of the work process—that is, individual differences in how respondents approached or completed the tasks—for instance (domain-specific) speed (Goldhammer and Klein Entink 2011; van der Linden 2007), the propensity to use a certain solution strategy (Greiff et al. 2016), or the use of planning when solving a complex problem (Eichmann et al. 2019).

The *evidence model,* (b), deals with the question of how to estimate the variables defined in the student model (constructs) given the observed performance of the respondent. For this purpose, two components are needed: the evidence rules and the measurement model. The *evidence rules* are used to identify observable evidence for the targeted construct. In this sense, log file data and process indicators calculated from it (see Sect. 10.1.2) provide evidence for assessing the process-related constructs mentioned above. For instance, the public use file of Round 1 of PIAAC (2011–2012) includes process indicators, such as the total time spent on an item, which can be used as an indicator of speed or for deriving indicators of test-taking engagement (Goldhammer et al. 2016). Log file data may also play an important role when identifying evidence for product-related constructs (e.g. ability, competence) measured by traditional product indicators. Here, log file data are a suitable complement to evidence rules in multiple ways. They can be used to obtain a more fine-grained (partial credit) scoring of the work product, depending on whether interactions contributing to the correct outcome were carried out or not (e.g. problem solving in PISA 2012; OECD 2013a), to inform the coding of missing responses (e.g. responses in PIAAC without any interaction and a time on task less than 5 s were coded as 'Not reached/not attempted'; OECD 2013b), and to detect suspicious cases showing aberrant response behaviour (van der Linden and Guo 2008) or data fabrication (Yamamoto and Lennon 2018).

As a second component, the evidence model includes a *statistical (measurement) model* for synthesising evidence across items. Here, multiple process indicators can identify a latent variable representing a process-related construct (e.g. planning, speed, test-taking engagement) and may complement product indicators to improve the construct representation. A more technical reason to identify a process-related construct is to make the estimation of the product-related (ability) construct more precise, which requires joint modelling of both constructs (e.g. two-dimensional ability–speed measurement models; Bolsinova and Tijmstra 2018; Klein Entink et

al. 2009). Related to that, timing data can be helpful to model the missing data mechanism (Pohl et al. 2019) and to investigate the comparability between modes (Kroehne et al. 2019). Another interesting application of process indicators within measurement models for ability constructs is to select the item response model that is appropriate for a particular observation, depending on the type of response behaviour (solution behaviour vs. rapid guessing; Wise and DeMars 2006).

The *task model,* (c), is about designing tasks and/or situations (i.e. item stimuli) in a way that the evidence required to infer the targeted student model variable is elicited. Regarding process indicators based on log file data, this means that item stimuli must provide adequate opportunities for interaction with the task environment (Goldhammer and Zehner 2017). In PIAAC, this issue is briefly discussed in the conceptual assessment framework for PS-TRE (PIAAC Expert Group in Problem Solving in Technology-Rich Environments 2009), where tasks require the respondent to operate and interact with (multiple) simulated software applications.

The *assembly model,* (d), refers to the combination of items on a test, determining how accurate the targeted construct is measured and how well the assessment represents the breadth of the construct. In adaptive testing, timing information can be used to improve item selection and thereby obtain a more efficient measurement (van der Linden 2008). Moreover, timing data can be used to optimise test design—in particular, to control the speededness of different test forms in adaptive testing (van der Linden 2005). In this elaborated sense, PIAAC (2011–2012) did not use timing information for its adaptive two-stage testing; however, timing information was used to assemble the cognitive assessment in order to obtain an expected overall test length of about 1 h for most of the respondents. An assessment may be adaptive or responsive not only in terms of item selection but also in a more general sense. Log file data may be used for triggering interventions if the response behaviour is not in line with the instruction—for example, if test-takers omit responses or switch to disengaged responding. This information can be fed back to the individual test-taker via prompts (Buerger et al. 2019), so that he or she can adapt, or to the proctor via a dashboard, so that he or she can intervene if needed (Wise et al. 2019).

10.1.2 Methodological Remarks: How to Identify States of Information Processing by Log File Data?

Log file data represent a type of paradata—that is, additional information about assessments generated as a by-product of computer-assisted data collection methods (Couper 1998). Extracting *process indicators* from log file data has not yet attracted much attention from a methodological perspective. The challenge becomes evident when one examines attempts to provide an overview of the heterogeneous types of paradata. The taxonomy of paradata provided by Kroehne and Goldhammer (2018), for example, shows that only a limited set of paradata can be directly

linked to the substantive data of the assessment (i.e. answers to questionnaire or test items). Only the response-related paradata—that is, all answer-change log events (e.g. selection of a radio button in a multiple choice item)—are directly related to the final response. However, the relationship of paradata to substantive data is of utmost interest when it comes to describing the test-taking process and explaining performance. Therefore, additional steps to process the information stored in log file data are necessary in order to extract meaningful *process indicators* that are related either at the surface (behavioural) level to the test-taking process or, preferably, to underlying cognitive processes (see Sect. 10.1.1).

Conceptually, the goal of creating process indicators can be described as the integration of three different sources of information: characteristics of the (evidence-centred) task design (see Sect. 10.1.1), expected (and observed) test-taking behaviour given the task design, and available log events specific to the concrete assessment system that may be suitable for inferring or reconstructing the test-taking behaviour. In combination with these sources, the process indicators are created to represent targeted attributes of the work process and to inform about the interaction between the test-taker and the task within an assessment platform (testing situation). At least two approaches can be conceptualised as to how process indicators can be defined: A first approach is to extract the indicators directly from the three aforementioned sources and to define them operationally with the concrete implementation in a particular programming language, such as Java, R, or SPSS syntax. A second more formal and generic approach can be distinguished, in which the algorithmic extraction of the indicators is first described and defined abstractly with respect to so-called states (Kroehne and Goldhammer 2018) before the actual process indicators are computed. In this framework, states are conceptualised as sections of the interaction between respondent and task within the assessment platform. How sections and states, respectively, are defined depends on the theory or model that is used to describe the test-taking and task solution process (e.g. a typical state would be 'reading the task instruction'). Indicators can be derived from properties of the reconstructed sequence of states (e.g. the total time spent on reading the task instruction). Reconstructing the sequence of states for a given test-taker from the log file data of a particular task requires that all transitions between states be identified with log events captured by the platform (e.g. by presenting the task instruction and the stimulus on different pages). Thus, the more formal approach to defining the extraction of process indicators from log file data provides the possibility of describing the relationship of the test-taking process to hypothesised cognitive processes (and their potential relationship to succeeding or failing in a task as represented by the substantive data). For that purpose, the theory-based mapping of states, transition between states, or sequences of state visits to cognitive processes is required.

However, the formal approach is not only relevant in the design phase of an assessment for planning, interpreting, and validating process indicators. It can also be used operationally to formally represent a given stream of events in a log file from the beginning to the end of a task. For that purpose, the log events are provided as input to one or multiple so-called finite state machines (Kroehne and

Goldhammer 2018) that process the events and change their state according to the machine's definition of states and state transitions triggered by certain events. This results in a reconstructed sequence of states for each test-taker who interacted with a particular task. Using properties of this reconstructed sequence of states allows for the extraction of process indicators, such as those programmed in the PIAAC LogDataAnalyzer (see Sect. 10.4.1, Table 10.2).

Given the available log file data, alternative states might be defined, depending on the specific research questions. The only requirement is the availability of log events that can be used to identify the transitions between states. If the log file data contain events that can be used to reconstruct the sequence of states for a particular decomposition of the test-taking process into states, indicators can be derived from properties of this reconstructed sequence.

The formal approach also allows the completeness of log file data to be judged. The general question of whether *all* log events are gathered in a specific assessment (e.g. PIAAC) is difficult or impossible to answer without considering the targeted process indicators. In this sense, log file completeness can be judged with respect to a known set of finite state machines representing all states and transitions of interest (e.g. to address a certain research question). If all transitions between states as defined by the finite state machine can be identified using information from the log file, the log file data are complete with respect to the finite state machines (described as *state completeness* in Kroehne and Goldhammer 2018).

10.2   Review of Research Work Using PIAAC Log File Data

With appropriate treatment, the PIAAC log file data (OECD 2017a, b, c, d, e, f, g, h, j, k, l, m, n, o, p, q, r) allow for the creation of a large number of informative indicators. Three generic process indicators derived from log file data are already included in the PIAAC Public Use File at the level of items—namely, total time on task, time to first action, and the number of interactions. This section provides a brief overview of the various research directions in which PIAAC log file data have been used so far. These studies include research on all three PIAAC domains, selected domains, and even specific items. They refer both to the data collected in the PIAAC main study and to the data collected in the field test, which was used to assemble the final instruments and to refine the operating procedures of the PIAAC main study (Kirsch et al. 2016). So far, PIAAC log files have been used to provide insights into the valid interpretation of test scores (e.g. Engelhardt and Goldhammer 2019; Goldhammer et al. 2014), test-taking engagement (e.g. Goldhammer et al. 2017a), dealing with missing responses (Weeks et al. 2016), and suspected data fabrication (Yamamoto and Lennon 2018). Other studies have concentrated on the highly interactive tasks in the domain of PS-TRE (He et al. 2019; He and von Davier 2015, 2016; Liao et al. 2019; Naumann et al. 2014; Stelter et al. 2015; Tóth et al. 2017; Vörös and Rouet 2016) and contributed to a better understanding of the adult competencies in operation.

The studies reviewed (Table 10.1) demonstrate how PIAAC log file data can contribute to describing the competencies of adults and the quality of test-taking, but as Maddox et al. (2018; see also Goldhammer and Zehner 2017) objected to the capturing of log events, inferences about cognitive processes are limited, and process indicators must be interpreted carefully.

10.2.1 Studies of Time Components Across Competence Domains

Processing times reflect the duration of cognitive processing when performing a task. Provided that information about the time allocation of individuals is available, several time-based indicators can be defined, such as the time until respondents first interact with a task or the time between the respondents' last action and their final response submission (OECD 2019). Previous research has often focused on 'time on task'—that is, the overall time that a respondent spent on the item. For example, analysis of the PIAAC log file data showed considerable variation of time on tasks in literacy and numeracy across countries, age groups, and levels of education, but comparatively less variability between the competence domains and gender (OECD 2019).

The (average) effect of time on task on a respondent's probability of task success is often referred to as 'time on task effect'. Using a mixed effect modelling approach, Goldhammer et al. (2014; see also Goldhammer et al. 2017a) found an overall positive relationship for the domain of problem solving, but a negative overall relationship for the domain of reading literacy. Based on theories of dual processing, this inverse pattern was explained in terms of different cognitive processes required; while problem solving requires a rather controlled processing of information, reading literacy relies on component skills that are highly automatised in skilled readers. The strength and direction of the time on task effect still varied according to individual skill level and task characteristics, such as the task difficulty and the type of tasks considered. Following this line of reasoning, Engelhardt and Goldhammer (2019) used a latent variable modelling approach to provide validity evidence for the construct interpretation of PIAAC literacy scores. They identified a latent speed factor based on the log-transformed time on task and demonstrated that the effect of reading speed on reading literacy becomes more positive for readers with highly automated word meaning activation skills, while—as hypothesised—no such positive interaction was revealed for perceptual speed.

Timing data are commonly used to derive indicators of disengagement (e.g. rapid guessing, rapid omissions) reflecting whether or not respondents have devoted sufficient effort to completing assigned tasks (Wise and Gao 2017). Several methods have been proposed that rely on response time thresholds, such as fixed thresholds (e.g. 3000 or 5000 ms) and visual inspection of the item-level response time distribution (for a brief description, see Goldhammer et al. 2016). The methods

Table 10.1 Overview of studies analysing log files from the PIAAC data base

Study	Data	Domain	Items(s)	Process representation
Engelhardt and Goldhammer (2019)	PIAAC-L (DE)	Literacy	49 items	Time on task
Goldhammer et al. (2014)	PIAAC field test (DE)	Literacy PS-TRE	49 literacy items 13 PS-TRE items	Time on task
Goldhammer et al. (2016)	PIAAC main study 22 countries	All	49 literacy items 49 numeracy items 14 PS-TRE items	Time on task
Goldhammer et al. (2017a)	PIAAC main study (CA)	All	49 literacy items 49 numeracy items 14 PS-TRE items	Time on task
He et al. (2019)	PIAAC main study (GB, IE, JP, NL, US)	PS-TRE	7 items (module PS2)	Longest common subsequences
He and von Davier (2015)	PIAAC main study (JP, NL, US)	PS-TRE	Club Membership Member ID	n-grams (unigrams, bigrams, trigrams)
He and von Davier (2016)	PIAAC main study (JP, NL, US)	PS-TRE	Club Membership Member ID	n-grams (unigrams, bigrams, trigrams)
Liao et al. (2019)	PIAAC main study (US)	PS-TRE	Meeting rooms	Time on task n-grams (unigrams, bigrams, trigrams)
Naumann et al. (2014)	PIAAC field test (CA, DE)	PS-TRE	20 items	No. of interactions
OECD (2019)	PIAAC main study 16 countries	All	49 literacy items 49 numeracy items 14 PS-TRE items	Time on task Time since last action Time to first interaction
Stelter et al. (2015)	PIAAC field test (DE)	PS-TRE	6 items	Time on task Time on routine steps
Tóth et al. (2017)	PIAAC field test (DE)	PS-TRE	*job search (released item)	No. of website visits No. of visits to different websites No. of visits to relevant websites Ratio of time spent on the relevant websites No. of bookmarked websites

(continued)

Table 10.1 (continued)

Vörös and Rouet (2016)	PIAAC main study 16 countries	PS-TRE	Club membership member ID, digital photography book purchase, meeting rooms	No. of interactions (recoded in categorical groups of low, medium, and high) Time on task (recoded in categorical groups of low, medium, and high)
Weeks et al. (2016)	PIAAC main study 22 countries	Literacy Numeracy	49 literacy items 49 numeracy items	Time on task
Yamamoto and Lennon (2018)	PIAAC main study 2 countries	All	BQ and cognitive assessment	Various aspects of time information, including time on task, time to first interaction, and timing associated with keystroke sequences

Note.   *https://piaac-logdata.tbahosting.de/public/problemsolving/JobSearchPart1/pages/jsp1-home.html

of P+ > 0% (Goldhammer et al. 2016, 2017a) and T-disengagement (OECD 2019) determine item-specific thresholds below which it is not assumed that respondents have made serious attempts to solve an item. P+ > 0% combines the response times with a probability level higher than that of a randomly correct response (in case of the PIAAC items, the chance level was assumed to be zero since most of the response formats allowed for a variety of different responses). The T-disengagement indicator further restricts this definition by implementing an additional 5-second boundary that treats all responses below this boundary as disengaged. Main results of these studies (Goldhammer et al. 2016, 2017a; OECD 2019) revealed that, although PIAAC is a low-stakes assessment, the proportions of disengagement across countries were comparatively low and consistent across domains. Nevertheless, disengagement rates differed significantly across countries, and the absolute level of disengagement was highest for the domain of problem solving. Other factors that promote disengagement included the respondents' level of education, the language in which the test was taken, respondents' level of proficiency, and their familiarity with ICT, as well as task characteristics, such as the difficulty and position of a task, which indicated a reduction in test-taking effort on more difficult tasks and tasks administered later in the assessment.

Similar to the issue of respondents' test engagement, time on task can be used to determine how to treat missing responses that may occur for various reasons, such as low ability, low motivation, or lack of time. In particular, omitted responses are an issue of the appropriate scaling of a test, because improperly treating omits as accidentally missing or incorrect could result in imprecise or biased estimates (Weeks et al. 2016). In the PIAAC main study (OECD 2013b), a missing response with no interaction and a response time under 5 s is treated as if the respondent did not see the item ('not reached/not attempted'). Timing information can help

to determine if this cut-off criterion is suitable and reflective of respondents having had enough time to respond to the item. Weeks et al. (2016) investigated the time on task associated with PIAAC respondents' assessment in literacy and numeracy to determine whether or not omitted responses should be treated as not administered or as incorrect. Based on descriptive results and model-based analyses comparing response times of incorrect and omitted responses, they concluded that the commonly used 5-second rule is suitable for the identification of rapidly given responses, whereas it would be too strict for assigning incorrect responses.

The consideration of time information was also used to detect data falsifications that can massively affect the comparability of results. Taking into account various aspects of time information, ranging from time on task to timing related to keystrokes, Yamamoto and Lennon (2018) argued that obtaining an identical series of responses is highly unlikely, especially considering PIAAC's adaptive multistage design. They described the cases of two countries that had attracted attention because a large number of respondents were interviewed by only a few interviewers. In these countries, the authors identified cases in which the processing of single cognitive modules was identical down to the time information; even entire cases were duplicated. Other results showed systematic omissions of cognitive modules with short response times. Consequently, suspicious cases (or parts of them) were dropped or treated as not administered in the corresponding countries.

10.2.2 Studies of the Domain of PS-TRE

The PIAAC domain of PS-TRE measures adult proficiency in dealing with problems related to the use of information and communication technologies (OECD 2012). Such problems can range from searching the web for suitable information to organising folder structures in digital environments. Accordingly, the PS-TRE tasks portray nonroutine settings requiring effective use of digital resources and the identification of necessary steps to access and process information. Within the PS-TRE tasks, cognitive processes of individuals and related sequences of states can be mapped onto explicit behavioural actions recorded during the problem-solving process. Clicks showing, for example, that a particular link or email has been accessed provide an indication of how and what information a person has collected. By contrast, other cognitive processes, such as evaluating the content of information, are more difficult to clearly associate with recorded events in log files.

Previous research in the domain of PS-TRE has analysed the relationship between problem-solving success and the way in which individuals interacted with the digital environment. They have drawn on a large number of methods and indicators for process analysis, which include the investigation of single indicators (e.g. Tóth et al. 2017) and entire action sequences (e.g. He and von Davier 2015, 2016).

A comparatively simple indicator that has a high predictive value for PS-TRE is the number of interactions with a digital environment during the problem-solving

process. Supporting the assumption that skilled problem solvers will engage in trial-and-error and exploration strategies, this action count positively predicted success in the PS-TRE tasks for the German and Canadian PIAAC field test data (Naumann et al. 2014; see also Goldhammer et al. 2017b) and for the 16 countries in the PIAAC main study (Vörös and Rouet 2016). Naumann et al. (2014) even found that the association was reversely U-shaped and moderated by the number of required steps in a task. Taking into account the time spent on PS-TRE tasks, Vörös and Rouet (2016) further showed that the overall positive relationship between the number of interactions and success on the PS-TRE tasks was constant across tasks, while the effect of time on task increased as a function of task difficulty. They also revealed different time–action patterns depending on task difficulty. Respondents who successfully completed an easy task were more likely to show either a low action count with a high time on task or a high action count with a low time on task. In contrast, the more time respondents spent on the task, and the more they interacted with it, the more likely they were to solve a medium and a hard task. Although both Naumann et al. (2014) and Vörös and Rouet (2016) investigated the respondents' interactions within the technology-rich environments, they used different operationalisations—namely, a log-transformed interaction count and a percentile grouping variable of low, medium, and high interaction counts, respectively. However, they obtained similar and even complementary results, indicating that the interpretation of interactions during the process of problem solving might be more complex than a more-is-better explanation, providing valuable information on solution behaviours and strategies.

Process indicators can also combine different process information. Stelter et al. (2015; see also Goldhammer et al. 2017b) investigated a log file indicator that combined the execution of particular steps in the PS-TRE tasks with time information. Assuming that a release of cognitive resources benefits the problem-solving process, they identified routine steps in six PS-TRE tasks (using a bookmark tool, moving an email, and closing a dialog box) and measured the time respondents needed to perform these steps by determining the time interval between events that started and ended sequences of interest (e.g. opening and closing the bookmark tool; see Sect. 10.4.2). By means of logistic regressions at the task level, they showed that the probability of success on the PS-TRE tasks tended to increase inversely with the time spent on routine steps, indicating that highly automated, routine processing supports the problem-solving process.

While the number of interactions and the time spent on routine steps are generic indicators applicable to several different tasks, indicators can also be highly task-specific. Tóth et al. (2017) classified the problem-solving behaviour of respondents of the German PIAAC field test using the data mining technique of decision trees. In the 'Job Search' item[1], which was included in the PIAAC field test and now serves as a released sample task of the PS-TRE domain, respondents were asked to

---

[1]https://piaac-logdata.tba-hosting.de/public/problemsolving/JobSearchPart1/pages/jsp1-home.html

bookmark websites of job search portals in a search engine environment that did not have a registration or fee requirement. The best predictors included as decision nodes were the number of different website visits (top node of the tree) and the number of bookmarked websites. Respondents who visited eight or more different websites and bookmarked exactly two websites had the highest chance of giving a correct response. Using this simple model, 96.7% of the respondents were correctly classified.

Other important contributions in analysing response behaviour in the domain of PS-TRE were made by adopting exploratory approaches from the field of text mining (He and von Davier 2015, 2016; Liao et al. 2019). He and von Davier (2015, 2016) detected and analysed robust n-grams—that is, sequences of n adjacent actions that were performed during the problem-solving process and have a high information value (e.g. the sequence [viewed_email_1, viewed_email_2, viewed_email_1] may represent a trigram of states suggesting that a respondent revisited the first email displayed after having seen the second email displayed). He and von Davier (2015, 2016) compared the frequencies of certain n-grams between persons who could solve a particular PS-TRE task and those who could not, as well as across three countries to determine which sequences were most common in these subgroups. The results were quite consistent across countries and showed that the high-performing group more often utilised search and sort tools and showed a clearer understanding of sub-goals compared to the low-performing group.

Similarly, Liao et al. (2019) detected typical action sequences for subgroups that were determined based on background variables, such as the monthly earnings (first vs. fourth quartile), level of educational attainment, age, test language, and skill use at work. They examined the action sequences generated within a task in which respondents were required to organise several meeting room requests using different digital environments including the web, a word processor, and an email interface. Findings by Liao et al. show not only which particular action sequences were most prominent in groups with different levels of background variables but also that the same log event might suggest different psychological interpretations depending on the subgroup. Transitions between the different digital environments, for example, may be an indication of undirected behaviour if they are the predominant feature; but they may also reflect steps necessary to accomplish the task if accompanied by a variety of other features. However, although such in-detail item analyses can provide deep insights into the respondents' processing, their results can hardly be generalised to other problem-solving tasks, as they are highly dependent on the analysed context.

Extending this research direction to a general perspective across multiple items, He et al. (2019) applied another method rooted in natural language processing and biostatistics by comparing entire action sequences of respondents with the optimal (partly multiple) solution paths of items. By doing so, they determined the longest common subsequence that the respondents' action sequences had in common with the optimal paths. He et al. were thus able to derive measurements on how similar the paths of the respondents were to the optimal sequence and how consistent they were between the items. They found that most respondents in the countries

investigated showed overall consistent behaviour patterns. More consistent patterns were observed in particularly good- and particularly poor-performing groups. A comparison of similarity across countries by items also showed the potential of the method for explaining why items might function differently between countries (differential item functioning, DIF; Holland and Wainer 1993), for instance, when an item is more difficult in one country than in the others.

10.3   The Released PIAAC Log File Data

With the aim of making the PIAAC log file data available to the research community and the public, the OECD provided funding for the development of infrastructure and software tools to disseminate and facilitate the use of the log file data. Carried out by the GESIS – Leibniz Institute for the Social Sciences and the DIPF | Leibniz Institute for Research and Information in Education, the PIAAC log file data were anonymised and archived; the log file data and the corresponding PIAAC tasks were described in interactive online documentation; and a software tool, the PIAAC LogDataAnalyzer (LDA), enables researchers to preprocess and analyse PIAAC log file data. In the following, the PIAAC test design is outlined (Sect. 10.3.1). With this background, the structure of the PIAAC log file data (Sect. 10.3.2) is presented, and we explain how the available documentation of items (see also Chap. 4 in this volume) and related log events can be used to make sense of the log file data (Sect. 10.3.3).

10.3.1   Overview of PIAAC Test Design

PIAAC (2011–2012) included several assessment parts and adaptive routing to ensure an efficient estimation of adult proficiencies in the target population (Fig. 10.1; for details, see Kirsch et al. 2016; OECD 2019; Chap. 2 in this volume). After a computer-assisted personal interview (CAPI), in which the background of respondents was surveyed (background questionnaire, BQ), the respondents started either with the computer-based assessment (CBA) by default or—if they did not report any experience with information and communication technologies (ICT)—with the paper-based assessment (PBA). When routed to the CBA, respondents were first asked to complete a core assessment in which their basic ICT skills and cognitive skills were assessed (CBA Core Stages 1 and 2). If they passed, they were led to the direct assessment of literacy, numeracy, or—in countries taking this option—problem solving in technology-rich environments (PS-TRE). On average, about 77% of the respondents in all the participating countries completed the direct assessment on the computer—for example, 82% in Germany and 84% in the United States (Mamedova and Pawlowski 2018).
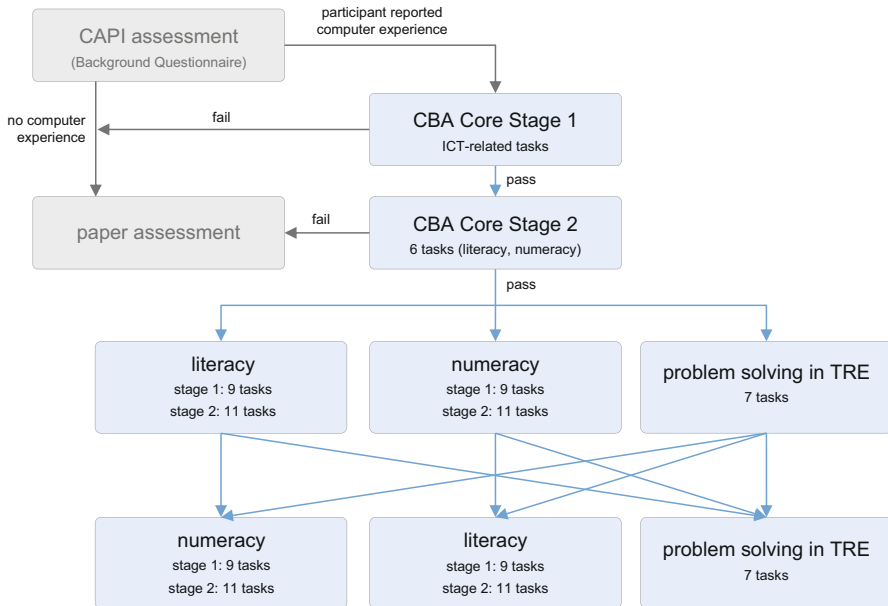
Fig. 10.1 Illustration of the CBA parts of the PIAAC assessment
Note. For the complete design, see OECD 2013b, p. 10

Terminologically, a set of cognitive items is called a *module*. Each respondent received two *modules* during the regular cognitive assessment. For the domains of literacy and numeracy, a module consists of two *testlets,* as literacy and numeracy were assessed adaptively. Specifically, both the literacy and the numeracy modules comprised two stages, each of which consisted of alternative testlets differing in difficulty (three testlets at Stage 1, four testlets at Stage 2). Note that within a stage, items could be included in more than one testlet. Due to the unique nature of PS-TRE, this domain was organised into two fixed modules of seven tasks each.

According to the conception and the design of the PIAAC study, the entire computer-based cognitive assessment was expected to take about 60 min. However, since PIAAC was not a timed assessment (OECD 2013b, p. 8), some respondents may have spent more time on the completion of the cognitive assessment. Log file data are available only for the CBA parts of the PIAAC study (coloured boxes in Fig. 10.1).

10.3.2 File Structure and Accessibility

The raw PIAAC log files are XML files that contain records of the respondents' interactions with the computer test application used for PIAAC (TAO: Jadoul et al.
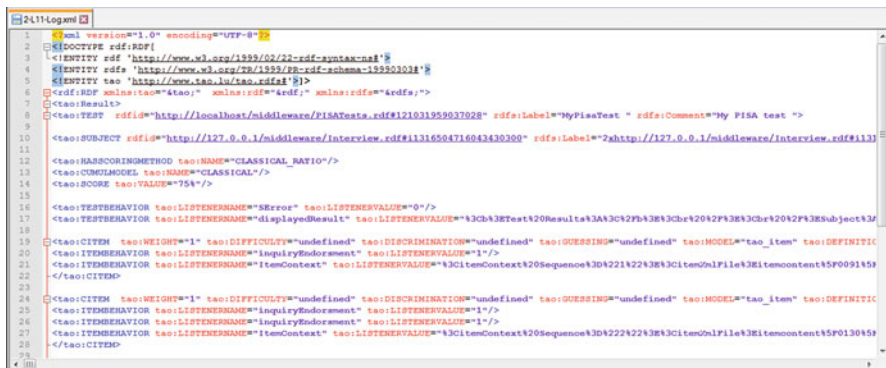
Fig. 10.2 Example of the appearance of a raw XML file

2016; CBA ItemBuilder: Rölke 2012). Specifically, the logged actions of respondents (e.g. starting a task, opening a website, selecting or typing an answer) were recorded and stored with time stamps. Figure 10.2 shows an example screenshot of the content of an XML file. However, users interested in working with the PIAAC log file data for literacy, numeracy, and PS-TRE items are not required to further process the raw XML files. Instead, they can use the PIAAC LogDataAnalyzer (LDA) as a tool for preprocessing and analysing the log file data (see Sect. 10.4; this does not apply to the log file data of the core assessment parts UIC and Core2).

There is a log file for each assessment component that a respondent took during the test. In total, 18 different components were administered, depending on the tested domain and the stage:

– CBA Core Stage 1 (UIC)
– CBA Core Stage 2 (Core2)
– Literacy at Stage 1 (three testlets, L11–L13) and Stage 2 (four testlets, L21–L24)
– Numeracy at Stage 1 (three testlets, N11–N13) and Stage 2 (four testlets, N21–N24)
– Problem solving in technology-rich environments (modules PS1 and PS2)

Figure 10.3 shows the log files of the respondent with the code 4747 as an example. This respondent completed the CBA Core assessments (UIC and Core2), two testlets of the literacy assessment (L12, L24), and two testlets of the numeracy assessment (N12, N24). Note that PS-TRE was an optional part of the cognitive assessment. For this reason, the data files of France, Italy, and Spain do not include any files with regard to PS-TRE.

As stated before, the log file data contain data only on the CBA parts of the PIAAC study. The responses to the background questionnaire (BQ) and the scored responses of the cognitive assessment are part of the PIAAC Public Use Files, which also include additional information, such as sampling weights, the results of the PBA, and observations of the interviewer. The public use files also include a limited set of process indicators (i.e. total time on task, time to first action, and the number

Name	Typ	Komprimie...	Kenn...	Größe	Verh...	Änderungsdatum
4747-Core2-Log	XML-Doku...	6 KB	Nein	73 KB	93%	25.10.2017 12:45
4747-L12-Log	XML-Doku...	10 KB	Nein	177 KB	95%	25.10.2017 12:45
4747-L24-Log	XML-Doku...	8 KB	Nein	108 KB	94%	25.10.2017 12:45
4747-N12-Log	XML-Doku...	6 KB	Nein	73 KB	93%	25.10.2017 12:45
4747-N24-Log	XML-Doku...	6 KB	Nein	83 KB	93%	25.10.2017 12:45
4747-UIC-Log	XML-Doku...	2 KB	Nein	16 KB	89%	25.10.2017 12:45

Fig. 10.3 Available XML raw files of the respondent with the code 4747

of interactions) for cognitive items of the CBA. The international PIAAC Public Use Files are available on the OECD website[2]. For academic research, the German Scientific Use File[3], including additional and more detailed variables than the international public use file, can be combined with the PIAAC log file data as well. The data can be merged using the variables CNTRYID, which is an identification key for each country, and SEQID, which is a unique identifier for individuals within each country.[4] By merging these data, detailed analyses can be carried out on, for example, how the behaviour of the respondents during the cognitive assessment relates to their task success or their background (see Sect. 10.2 for examples).

Researchers and academic teachers who wish to work with the PIAAC log file data (OECD 2017a – OECD 2017r) must register with the GESIS Data Archive[5] and the data repository service datorium (see also Chap. 4 in this volume). Accepting the terms of use will provide access to the PIAAC log file data[6] under the study code 'ZA6712, Programme for the International Assessment of Adult Competencies (PIAAC), log files', which then allows access to the raw PIAAC log files, compressed in ZIP files per country. All downloadable files have been anonymised—that is, all information that potentially identifies individual respondents has been removed or replaced with neutral character strings. Otherwise, the log file data are as complete as they were when logged for individual respondents. In addition to data access, the GESIS Data Archive provides users of PIAAC log file data with further information—for example, on the bibliographic citation and descriptions of content and methodological aspects of the PIAAC log file data.[7]

---

[2]http://www.oecd.org/skills/piaac/publicdataandanalysis/

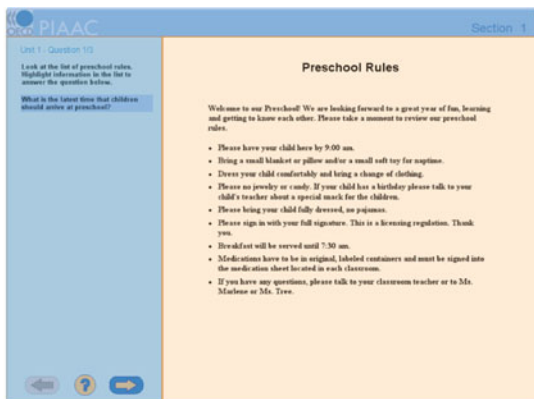[3]https://www.gesis.org/en/piaac/rdc/data/national-scientific-use-files

[4]Note that SEQID is not unique across countries and therefore has to be combined with CNTRYID to create unique individual identifiers across countries.

[5]https://search.gesis.org/research_data/ZA6712

[6]https://doi.org/10.4232/1.12955

[7]see also: https://www.gesis.org/en/piaac/rdc/data/international-piaac-log-files

Fig. 10.4 Top: A PIAAC item consists of a question part including general instructions (blue-shaded) and a stimulus part (orange-shaded). Bottom: Example of recorded events in a data frame format for a respondent working on the first item of the testlet L13

Notes. See also Sect. 10.4.2; Log events that reflect interactions with the question part are blue-shaded (event_name: *taoPIAAC*); log events that reflect interactions with the stimulus are orange-shaded (event_name: *stimulus*)

10.3.3 Documentation of Items and Log Events

Before using the PIAAC log file data, users should be aware that PIAAC items consist of two general parts—the question part including general instructions on the blue background (*taoPIAAC*) and the stimulus next to it (*stimulus*; Fig. 10.4, top). The elements in the PIAAC log file data are assigned to these parts (Fig. 10.4, bottom). Depending on the item response format, respondents were required to give a response using elements of the question part (e.g. entering input in a text field on the left panel) or the stimulus (e.g. highlighting text in the stimulus). An overview of the specific response formats of the literacy and numeracy items can be retrieved from the online PIAAC Log Data Documentation.[8]

---

[8] https://piaac-logdata.tba-hosting.de/itemid.html

The PIAAC items are documented in the PIAAC Reader's Companion[9] as well as in the online PIAAC Log Data Documentation.[10] While the Reader's Companion briefly summarises the assessed competence domains and gives a general description of the items in the cognitive assessment and the background questionnaire, the online documentation displays the exact items and interactively provides details of the mapping with events in the PIAAC log file data (Fig. 10.5). The documentation of released items is available for all users (e.g. Job Search,[11] MP3 Player[12]). Although they were not administered as part of the PIAAC main study, the released items demonstrate how PIAAC items are documented. If researchers wish to access the full PIAAC item documentation and the items of the main study, they must complete an application form[13] including a short description of their research interest and a signed confidentiality agreement and send it to the contact officer at the OECD.[14] In case of a successful application, researchers will receive a username and password for the online platform with which they can access all documentation.

In the online documentation, possible event types are displayed in the form of pop-up dialogs where they occur within the items. The pop-up dialogs are activated when the mouse cursor moves over a sensitive item element. The documentation includes all items of the PS-TRE domain and a subset of the literacy and numeracy items that demonstrate the implemented response formats and therefore represent the range of possible log events in literacy and numeracy items. The logged events in the generated XML files follow a particular structure:

```
<taoEvent Name ="origin" Type ="Event Type" Time ="ms">
                                            </taoEvent>
```

An event tag bracketed by `<taoEvent>` and `</taoEvent>` denotes that one interaction of the respondent with the item environment was recorded. An interaction with one item element might trigger the logging of multiple event tags. The attributes within the tags specify the interaction in detail. The attribute `Name` states the environment of the element with which a respondent interacted (e.g. *taoPIAAC* or *stimulus*). The attribute `Type` classifies the recorded event (e.g. TOOLBAR, MENU, or BUTTON), while the attribute `Time` provides a time stamp in milliseconds, which is reset to zero at the start of each unit. A list[15] of all possible

---

[9]https://www.oecd.org/skills/piaac/Skills%20(vol%202)-Reader%20companion%2D%2Dv7%20eBook%20(Press%20quality)-29%20oct%200213.pdf

[10]https://piaac-logdata.tba-hosting.de/

[11]https://piaac-logdata.tba-hosting.de/public/problemsolving/JobSearchPart1/pages/jsp1-home.html

[12]https://piaac-logdata.tba-hosting.de/public/problemsolving/MP3/pages/mp3-start.html

[13]http://www.oecd.org/skills/piaac/log-file/Application.pdf

[14]edu.piaac@oecd.org

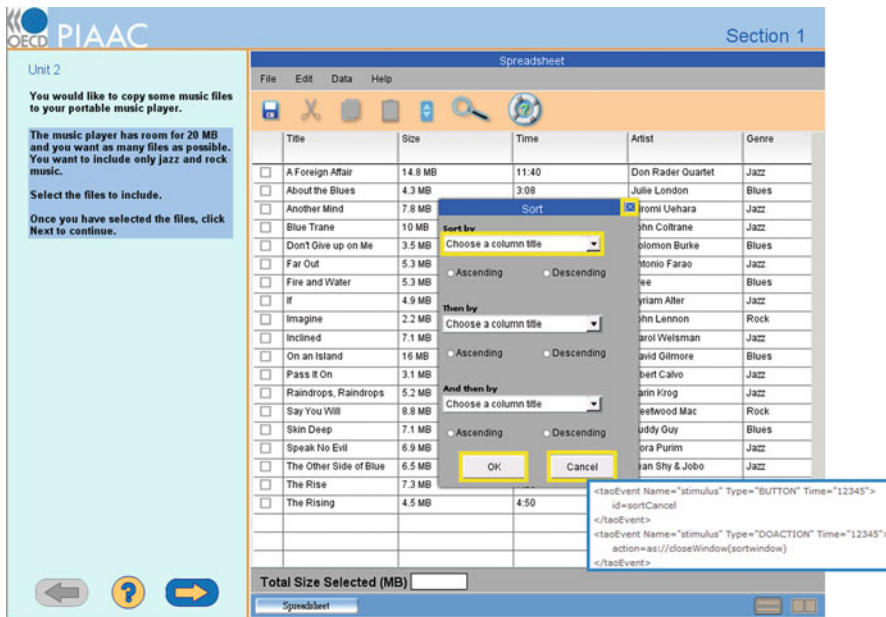[15]https://piaac-logdata.tba-hosting.de/additionalinformation.html

Fig. 10.5 Example item
Notes. MP3 Player (see also OECD 2012, *Literacy, Numeracy and Problem Solving in Technology-Rich Environments*, p. 55.). Moving the mouse cursor over sensitive areas (here the Cancel button) displays blue-framed pop-up dialogs containing details about the structure of the recorded events. Yellow-framed areas are clickable parts of the item documentation and open new screens. Available at: https://piaac-logdata.tba-hosting.de/public/problemsolving/MP3/pages/mp3-start.html

log events is provided in the online PIAAC Log Data Documentation for the items assessing literacy and numeracy (tab *Events-Literacy, Numeracy*) and PS-TRE (tab *Events-Problem Solving in Technology-Rich Environments*)[16].

10.4   Preprocessing and Analysing PIAAC Log File Data

The software PIAAC LogDataAnalyzer (LDA) was developed for the OECD by the TBA Centre at the DIPF in cooperation with TBA21 Hungary Limited to facilitate the analysis of log file data—that is, to handle the huge amount of XML files (e.g. the log file data from Germany comprises 24.320 XML files of 1.9 gigabytes) and to preprocess the log file data for further analyses in statistical software packages, such as R (R Core Team 2016).

---

[16]https://piaac-logdata.tba-hosting.de/
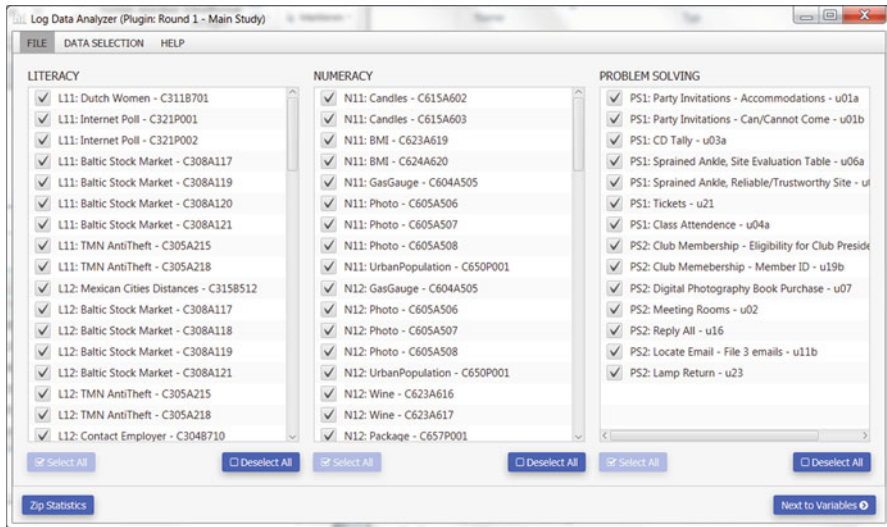
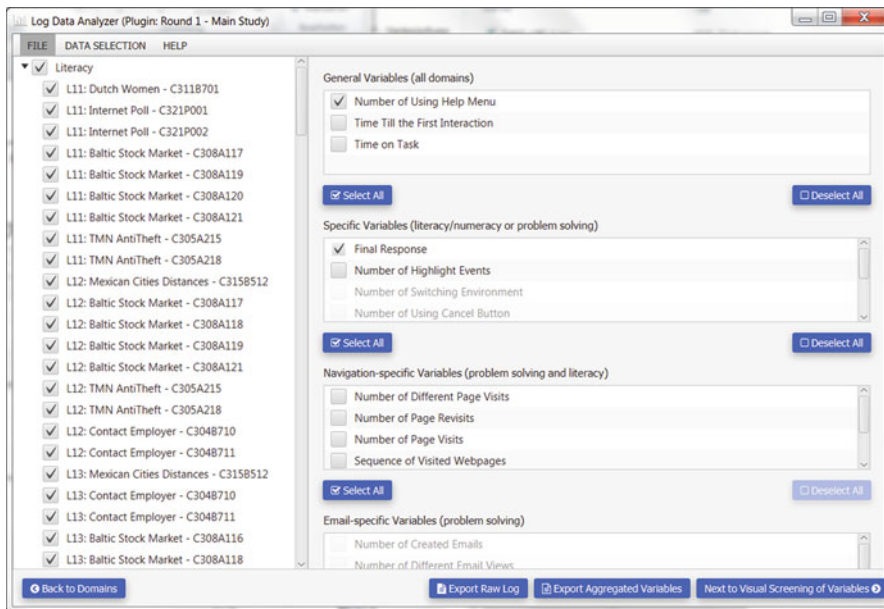**Fig. 10.6** PIAAC LDA user interface for the selection of items by domain

Essentially, the PIAAC LDA fulfils two main purposes: the extraction of predefined aggregated variables from XML files (Sect. 10.4.1) and the extraction of raw log file data from XML files for creating user-defined aggregated variables (Sect. 10.4.2). Note that all the XML files from the literacy, numeracy, and problem-solving modules are included in the preprocessing, but the XML files from CBA Core Stage 1 (UIC) and CBA Core Stage 2 (Core2) are not.

The LDA software was developed for MS Windows (e.g. 7, 10; both 32 bit and 64 bit) and can be accessed and downloaded via the OECD's PIAAC Log File Website[17]. The help area of the LDA does include detailed information about the LDA software itself. It describes how to import a ZIP file that includes a country's log file data; how to select items within domains; how to select aggregated variables (including visual screening); and how to export aggregated variables and raw log file data. Furthermore, there is extensive information about how the LDA software handles errors or unexpected issues (e.g. the handling of negative time stamps).

The typical workflow for using the PIAAC LDA starts with importing a country's ZIP file, including all the XML files. For demonstration purposes, the LDA also provides a small ZIP file (sample_round1_main.ZIP). The import ends with a brief report presenting details about the ZIP file and the included data.[18] Next, the user can select the domains and items by selecting those of interest and deselecting those to be excluded (see Fig. 10.6).

---

[17]http://piaac-logdata.tba-hosting.de/download/

[18]If the verification of the ZIP file failed, for instance, because a corrupt or wrong ZIP file was selected, an error message is presented that the data is invalid.

Fig. 10.7 PIAAC LDA user interface for exporting aggregated variables and raw log file data

This selection should be used if only a subset of data is needed for a particular research question. All subsequent data processing steps include only this subset, which helps to reduce data processing time. Note that due to the PIAAC test design (see Sect. 10.3.1), a particular item may be included in more than one testlet (e.g. Photo—C605A506 is part of testlet N11 and testlet N12; see Fig. 10.6). To gather the complete data for an item, all occurrences in testlets have to be selected. An overview of the booklet structure, including testlets, can be found in the online PIAAC Log Data Documentation (tab *Booklet Order*).[19]

Clicking the button 'Next to Variables' (Fig. 10.6) opens the view for exporting data for the selected items (Fig. 10.7). If needed, the item selection can be modified in the left panel.

10.4.1   *Aggregated Variables Provided by the PIAAC Log Data Analyzer*

The selection of aggregated variables offered in the right panel depends on the selection of items, because some of the aggregated variables are available only for

---

[19]https://piaac-logdata.tba-hosting.de/additionalinformation.html

specific items. The set of aggregated variables offered by the LDA was predefined by subject matter experts on the assumption that these variables are highly relevant for researchers interested in the analysis of log file data. Table 10.2 gives an overview of the available aggregated variables that apply either to all items (*general*), to selected items depending on the domain (*item-specific*), or to selected items depending on the item content, such as simulated websites in literacy and problem solving (*navigation-specific*) or simulated email applications in PS-TRE (*email-specific*).

Following Sect. 10.1.2, the indicators provided by the LDA can be represented in terms of properties of the reconstructed sequence of states (Table 10.2, columns *Defined State(s)* and *Property of the reconstructed sequence*). Describing the indicators provided by the LDA with respect to states used in the more formal approach (see Sect. 10.1.2) allows for the similarities between the indicators to become apparent.

A detailed description of all the aggregated variables can be found in the help menu of the PIAAC LDA. Figure 10.8 shows an example for the variable 'Final Response' given that the response mode is 'Stimulus Clicking'. The help page provides some general information about the variable and how it is extracted from log events (e.g. event 'itemScoreResult' including the attribute 'selectedImageAreas'). This is complemented by an example log and the final response as extracted from the log (e.g. final response: I3|I4, indicating that the image areas I3 and I4 were selected by the respondent).

By pressing 'Next to Visual Screening of Variables' (Fig. 10.7), the PIAAC LDA offers a screening for aggregated variables. The user can quickly examine the selected variables by generating simple descriptive statistics (minimum, maximum, average, and standard deviation) and charts (pie chart or histogram). Based on the screening, for instance, variables showing no variance could be deselected.

The export of selected aggregated variables is done by pressing the button 'Export Aggregated Variables'. The output file is a CSV file using tabulators as separators. The traditional wide format shows cases in rows and variables in columns. The first two columns show the CNTRYID and SEQID, which are needed to merge aggregated variables with variables from the PIAAC Public Use File. The CSV file can be imported by statistical analysis software, such as R, for further investigation.

10.4.2   User-Defined Aggregated Variables

Researchers who cannot find what they are looking for in the set of predefined variables can export the raw log file data of selected items by pressing the 'Export Raw Log' button. Based on this exported data, they can create their own aggregated variables using statistical software. The output file is a text file that includes all the log events stored in the XML files. The output file has a long format—that is, the log events of a person extend over multiple rows as indicated by identical CNTRYID and SEQID (Fig. 10.9). Each line in the file contains information that

Table 10.2 Aggregated variables provided by the PIAAC LDA

Indicator created by the PIAAC LogDataAnalyzer (LDA)	Type	Domain			Defined state(s)	Property of the reconstructed sequence
		L	N	PS		
Time on task	G	x	x	x	One single state corresponding to the complete task	Total time on the state
Number of using cancel button	S	–	–	x		Count of a dedicated event
Number of using help	G	x	x	x		Count of a dedicated event
Number of highlight events	S	x	–	–		Count of a dedicated event
Time till the first interaction	G	x	x	x	Three states for each task (one prior to the first interaction, one after the last answer selection, one in between)	Total time on the first state
Time since last answer interaction	S	x	x	–		Total time on the third state
Number of switching environment	S	–	–	x^1	State for each environment (in relevant tasks)	Number of transitions
Sequence of switching environment	S	–	–	x^1		Sequence of states (n-grams)
Number of created email	E	–	–	x^2	State for the email environment (in relevant tasks)	Count of dedicated events
Sequence of viewed email	E	–	–	x^2	State for each email (in relevant tasks)	Sequence of states (n-grams)
Number of email views	E	–	–	x^2		Number of state visits
Number of different email views	E	–	–	x^2		Number of unique state visits
Number of revisited emails	E	–	–	x^2		Number of state revisits
Sequence of visited webpages	N	x^3	–	x^3	States for each page (in relevant tasks)	Sequence of states (n-grams)
Time sequence of spent time on webpages	N	x^3	–	x^3		Sequence of time on state for a state sequence (n-grams)
Number of different page visits	N	x^3	–	x^3		Number of unique state visits
Number of page visits	N	x^3	–	x^3		Number of state visits
Number of page revisits	N	x^3	–	x^3		Number of state revisits

Notes. [1]Available for items including multiple software applications; [2]Available for items including a simulated email application; [3]Available for items including a simulated web browser; Domain: *L* literacy, *N* numeracy, *PS* problem solving; Type: *G* general variables, *S* specific variables, *E* email-specific variables, *N* navigation-specific variables
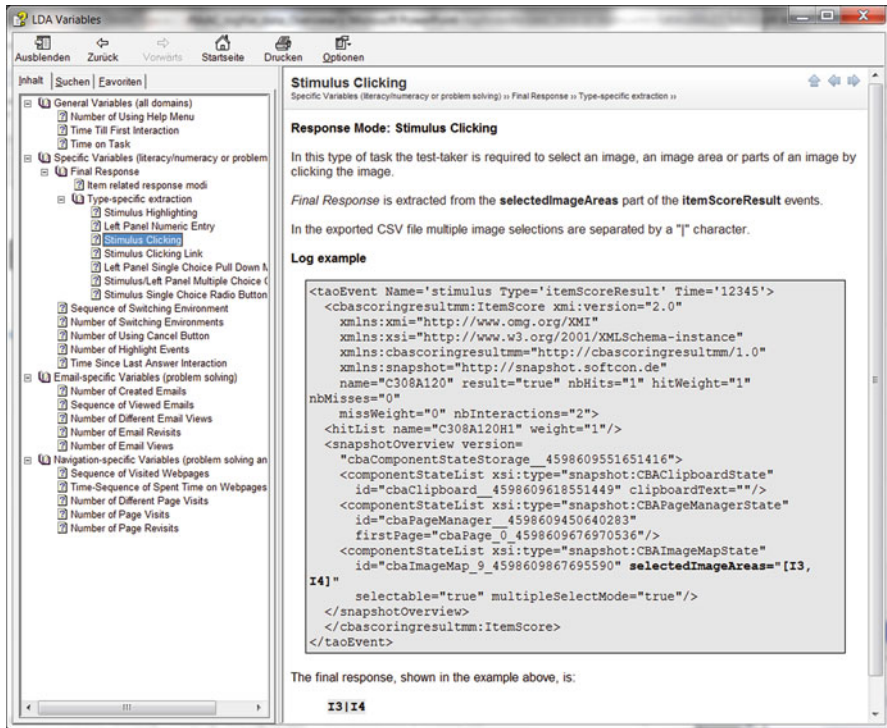
Fig. 10.8 Example help page from the PIAAC LogDataAnalyzer (LDA) providing information about the aggregated variable 'Final response' for the response mode 'Stimulus Clicking'

was extracted from the attributes of the event tags in the XML log files (see Sect. 10.3.3). This information is converted into a data table format separated by tabs with the following columns: CNTRYID, SEQID, booklet_id, item_id, event_name, event_type, time_stamp, and event_description.

The booklet_id indicates the testlet for literacy and numeracy and the module for PS-TRE. The item_id gives the position of the item in the respective testlet or module. The corresponding PIAAC item identifier can be obtained from the booklet documentation (see the online PIAAC Log Data Documentation[20], tab *Booklet Order*). The event_name represents the environment from which the event is logged. Values can be *stimulus* for events traced within the stimulus, *taoPIAAC* for events from outside the stimulus, and *service* for other parts of a unit. The event_type represents the event category. The time_stamp provides a time stamp in milliseconds with the beginning of each unit as reference point (time_stamp: 0). Finally, the event_description provides the values that characterise a certain event

---

[20] https://piaac-logdata.tba-hosting.de/additionalinformation.html

CNTRYID	SEQID	booklet_id	item_id	event_name	event_type	timestamp	event_description
...	...	...	...	...	...	...	
1 sample_round1_main	295	PS1	5	taoPIAAC	START	0	TEST_TIME=327
2 sample_round1_main	295	PS1	5	stimulus	MENU	4050	id=file-menu
3 sample_round1_main	295	PS1	5	stimulus	MENUITEM	7315	key=cut
4 sample_round1_main	295	PS1	5	stimulus	COPY	7315	content=undefined
5 sample_round1_main	295	PS1	5	stimulus	TOOLBAR	17184	id=toolbar_back_btn
6 sample_round1_main	295	PS1	5	stimulus	HISTORY_BACK	17184	pageId=unit06bdefault|*$title=Websearch\|*$url=www.websearch.com/treatmentsprainedankle
7 sample_round1_main	295	PS1	5	stimulus	DOACTION	17187	action=as://historyBack(toolbar_back_btn)
8 sample_round1_main	295	PS1	5	stimulus	TOOLBAR	19222	id=toolbar_forward_btn
9 sample_round1_main	295	PS1	5	stimulus	HISTORY_NEXT	19222	pageId=unit06bdefault|*$title=Websearch\|*$url=www.websearch.com/treatmentsprainedankle
10 sample_round1_main	295	PS1	5	stimulus	DOACTION	19224	action=as://historyForward(toolbar_forward_btn)
11 sample_round1_main	295	PS1	5	stimulus	TOOLBAR	21906	id=toolbar_home_btn
12 sample_round1_main	295	PS1	5	stimulus	HISTORY_ADD	21909	pageId=unit06bdefault|*$title=Websearch\|*$url=www.websearch.com/treatmentsprainedankle
13 sample_round1_main	295	PS1	5	stimulus	DOACTION	21909	action=as://gotoURL(unit06bdefault,_self,unknown)
14 sample_round1_main	295	PS1	5	stimulus	MENU	24954	id=wb-bookmarks-menu
15 sample_round1_main	295	PS1	5	stimulus	MENUITEM	26987	key=bookmark-add
16 sample_round1_main	295	PS1	5	stimulus	BOOKMARK_ADD	30772	pageId=unit06bdefault|*$title=Websearch\|*$url=www.websearch.com/treatmentsprainedankle
17 sample_round1_main	295	PS1	5	stimulus	BUTTON	30977	id=add_bookmark_validation
18 sample_round1_main	295	PS1	5	stimulus	DOACTION	30977	action=as://hide(addbookmark_window)
19 sample_round1_main	295	PS1	5	stimulus	MENU	33677	id=wb-bookmarks-menu
20 sample_round1_main	295	PS1	5	stimulus	MENUITEM	34842	key=bookmark-manage
21 sample_round1_main	295	PS1	5	stimulus	BUTTON	40915	id=manage_bookmark_window_validation
22 sample_round1_main	295	PS1	5	stimulus	DOACTION	40916	action=as://hide(managebookmark_window)
23 sample_round1_main	295	PS1	5	stimulus	TEXTBOX_ONFO	45326	id=u06b_default_txt2|*$value=Treatment sprained ankle
...	...	...	...	...			

Fig. 10.9 Example raw log file data extracted from XML files

type. A detailed overview of possible event types and related values can be found in the online PIAAC Log Data Documentation.[21]

Figure 10.9 shows an example of raw log file data extracted from XML files and transformed to a long format. The test-taker is identified by CNTRYID (CNTRYID=sample_round1_main) and SEQID (SEQID: 295). He or she completed the fifth item (item_id: 5, which is "U06b - Sprained Ankle") in the PS-TRE module PS1 (booklet_id: PS1). Line 14 in Fig. 10.9 shows that the test-taker needed 24,954 milliseconds after starting the unit (time_stamp: 24954) to click on the menu bar of a simulated web browser (event_type: MENU) within the stimulus (event_name: stimulus); more specifically, it was the menu for bookmarking (event_description: id=wb-bookmarks-menu).

New aggregated variables can be extracted from this exported raw log file data. For example, a researcher might be interested in the time a test-taker needs to bookmark the current webpage (Stelter et al. 2015). Such an indicator can be described as the time in a particular state and understood as a specific part of the interaction between the test-taker and assessment platform (i.e. item), which, here, would be the of bookmarking. Traces of the test-taking process in the log file data can be used to reconstruct the occurrence of this state by using the menu events 'clicking the bookmarking menu', which describes the beginning of the state (event_description: id = wb-bookmarks-menu; line 14 in Fig. 10.9) and 'confirm the intended bookmark', which describes the end of the state and the transition to any subsequent state (event_description: id=add_bookmark_validation; line 17 in Fig. 10.9). When the two identifying events can be determined without considering previous or subsequent events, as in this example, indicators can be extracted directly from the log file data without specific tools. Instead, it is sufficient to filter the raw log events in such a way that only the two events of interest remain in chronological order in a dataset. The value for the indicator per person and for each

---

[21] https://piaac-logdata.tba-hosting.de/additionalinformation.html

state occurrence is the difference of the time stamps of successive events. Using the example provided in Fig. 10.9, the value of the indicator for that particular respondent is the difference of the two time stamps in lines 17 and 14 (i.e. 30,977–24,954 = 6023 milliseconds). If the meaning of events requires the context of the previous and subsequent events to be taken into account, algorithms to extract process indicators from log files can be formulated—for example, using finite state machines as described by Kroehne and Goldhammer (2018).

10.5   Conclusions

A major goal of this chapter was to make the PIAAC log file data accessible to researchers. As demonstrated for the case of PIAAC, providing this kind of data has some implications and challenges. It requires dealing with data formats and data structures other than usual assessment data while no established standards are yet available. As a consequence, there is also a lack of general tools that facilitate the preprocessing and transformation of raw log events. Another major issue is the documentation of items and log events triggered by the respondent's interaction with the assessment system. Without proper documentation, researchers who were not involved in the item development process can hardly make sense of the raw log data and therefore will not use it.

Assuming that a supporting infrastructure is available in the future, the use of log file data will likely no longer be an exception; rather, their use will grow in popularity. However, as indicated above, the creation of meaningful process indicators is limited in that they can be inferred only from states for which a beginning and an end are identified in the log data. This depends conceptually on the careful design of the task and its interactivity and technically on the design of the assessment system. Another issue to be considered is the interpretation of process indicators, which needs to be challenged by appropriate validation strategies (as it is usually required for the interpretation of test scores). Moreover, opening the stage for process indicators also requires statistical models to appropriately capture the more complex structure of dependencies between process and product indicators within items (e.g. Goldhammer et al. 2014; Klotzke and Fox 2019). Overall, users of log file data will have to face a series of conceptual and methodological tasks and challenges, but they will also be able to gain deeper insights into the behaviour and information processing of respondents.

References

Bolsinova, M., & Tijmstra, J. (2018). Improving precision of ability estimation: Getting more from response times. *British Journal of Mathematical and Statistical Psychology, 71*, 13–38. https://doi.org/10.1111/bmsp.12104.

Buerger, S., Kroehne, U., Koehler, C., & Goldhammer, F. (2019). What makes the difference? The impact of item properties on mode effects in reading assessments. *Studies in Educational Evaluation, 62*, 1–9. https://doi.org/10.1016/j.stueduc.2019.04.005.

Couper, M. (1998). Measuring survey quality in a CASIC environment. In: *Proceedings of the section on survey research methods of the American Statistical Association*, (pp. 41–49).

Eichmann, B., Goldhammer, F., Greiff, S., Pucite, L., & Naumann, J. (2019). The role of planning in complex problem solving. *Computers & Education, 128*, 1–12. https://doi.org/10.1016/j.compedu.2018.08.004.

Engelhardt, L., & Goldhammer, F. (2019). Validating test score interpretations using time information. *Frontiers in Psychology, 10*, 1–30.

Goldhammer, F., & Klein Entink, R. H. (2011). Speed of reasoning and its relation to reasoning ability. *Intelligence, 39*, 108–119. https://doi.org/10.1016/j.intell.2011.02.001.

Goldhammer, F., & Zehner, F. (2017). What to make of and how to interpret process data. *Measurement: Interdisciplinary Research and Perspectives, 15*, 128–132. https://doi.org/10.1080/15366367.2017.1411651.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*, 608–626.

Goldhammer, F., Martens, T., Christoph, G., & Lüdtke, O. (2016). *Test-taking engagement in PIAAC* (no. 133). Paris: OECD Publishing. https://doi.org/10.1787/5jlzfl6fhxs2-en.

Goldhammer, F., Martens, T., & Lüdtke, O. (2017a). Conditioning factors of test-taking engagement in PIAAC: An exploratory IRT modelling approach considering person and item characteristics. *Large-Scale Assessments in Education, 5*, 1–25. https://doi.org/10.1186/s40536-017-0051-9.

Goldhammer, F., Naumann, J., Rölke, H., Stelter, A., & Tóth, K. (2017b). Relating product data to process data from computer-based competency assessment. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education* (pp. 407–425). Cham: Springer. https://doi.org/10.1007/978-3-319-50030-0_24.

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior, 61*, 36–46. https://doi.org/10.1016/j.chb.2016.02.095.

He, Q., & von Davier, M. (2015). Identifying feature sequences from process data in problem-solving items with n-grams. In L. A. van der Ark, D. M. Bolt, W.-C. Wang, J. A. Douglas, & S.-M. Chow (Eds.), *Quantitative psychology research* (Vol. 140, pp. 173–190). Cham: Springer. https://doi.org/10.1007/978-3-319-19977-1_13.

He, Q., & von Davier, M. (2016). Analyzing process data from problem-solving items with n-grams: Insights from a computer-based large-scale assessment. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 749–776). Hershey: IGI Global. https://doi.org/10.4018/978-1-4666-9441-5.

He, Q., Borgonovi, F., & Paccagnella, M. (2019). *Using process data to identify generalized patterns across problem-solving items*. Paper presented at the Annual Meeting of National Council on Measurement in Education, Toronto, CA.

Holland, P., & Wainer, H. (Eds.). (1993). Differential item functioning. New York: Routledge. https://doi.org/10.4324/9780203357811.

Jadoul, R., Plichart, P., Bogaerts, J., Henry, C., & Latour, T. (2016). The TAO platform. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC) (2nd edition)*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf

Kirsch, I., Yamamoto, K., & Garber, D. (2016). Assessment design. In OECD (Ed.), *Technical report of the Survey of Adult Skills (PIAAC) (2nd edition)*. Paris: OECD Publishing. Retrieved from https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf

Klein Entink, R. H., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika, 74*, 21–48. https://doi.org/10.1007/s11336-008-9075-y.

Klotzke, K., & Fox, J.-P. (2019). Bayesian covariance structure modeling of responses and process data. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.01675.

Kroehne, U., & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika, 45*, 527–563. https://doi.org/10.1007/s41237-018-0063-y.

Kroehne, U., Hahnel, C., & Goldhammer, F. (2019). Invariance of the response processes between gender and modes in an assessment of reading. *Frontiers in Applied Mathematics and Statistics, 5*, 1–16. https://doi.org/10.3389/fams.2019.00002.

Liao, D., He, Q., & Jiao, H. (2019). Mapping background variables with sequential patterns in problem-solving environments: An investigation of United States adults' employment status in PIAAC. *Frontiers in Psychology, 10*, 1–32. https://doi.org/10.3389/fpsyg.2019.00646.

Maddox, B., Bayliss, A. P., Fleming, P., Engelhardt, P. E., Edwards, S. G., & Borgonovi, F. (2018). Observing response processes with eye tracking in international large-scale assessments: Evidence from the OECD PIAAC assessment. *European Journal of Psychology of Education, 33*, 543–558. https://doi.org/10.1007/s10212-018-0380-2.

Mamedova, S., & Pawlowski, E. (2018). *Statistics in brief: A description of U.S. adults who are not digitally literate*. National Center for Education Statistics (NCES). Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018161

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (ETS research report series, 1–29). https://doi.org/10.1002/j.2333-8504.2003.tb01908.x.

Naumann, J., Goldhammer, F., Rölke, H., & Stelter, A. (2014). Erfolgreiches Problemlösen in technologiebasierten Umgebungen: Wechselwirkungen zwischen Interaktionsschritten und Aufgabenanforderungen [Successful Problem Solving in Technology Rich Environments: Interactions Between Number of Actions and Task Demands]. *Zeitschrift für Pädagogische Psychologie, 28*, 193–203. https://doi.org/10.1024/1010-0652/a000134.

Organisation for Economic Co-operation and Development (OECD). (2019). *Beyond proficiency: Using log files to understand respondent behaviour in the survey of adult skills*. Paris: OECD Publishing. https://doi.org/10.1787/0b1414ed-en.

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*. Paris: OECD Publishing. https://doi.org/10.1787/9789264128859-en.

Organisation for Economic Co-operation and Development (OECD). (2013a). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2013b). *Technical report of the survey of adult skills (PIAAC)*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). (2017a). *Programme for the International Assessment of Adult Competencies (PIAAC), Austria log file*. Data file version 2.0.0 [ZA6712_AT.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017b). *Programme for the International Assessment of Adult Competencies (PIAAC), Belgium log file*. Data file version 2.0.0 [ZA6712_BE.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017c). *Programme for the International Assessment of Adult Competencies (PIAAC), Germany log file*. Data file version 2.0.0 [ZA6712_DE.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017d). P*rogramme for the International Assessment of Adult Competencies (PIAAC), Denmark log file*. Data file version 2.0.0 [ZA6712_DK.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017e). *Programme for the International Assessment of Adult Competencies (PIAAC), Estonia log file*. Data file version 2.0.0 [ZA6712_EE.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017f). *Programme for the International Assessment of Adult Competencies (PIAAC), Spain log file*. Data file version 2.0.0 [ZA6712_ES.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017g). *Programme for the International Assessment of Adult Competencies (PIAAC), Finland log file*. Data file version 2.0.0 [ZA6712_FI.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017h). *Programme for the International Assessment of Adult Competencies (PIAAC), France log file*. Data file version 2.0.0 [ZA6712_FR.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017j). *Programme for the International Assessment of Adult Competencies (PIAAC), United Kingdom log file*. Data file version 2.0.0 [ZA6712_GB.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017k). *Programme for the International Assessment of Adult Competencies (PIAAC), Ireland log file*. Data file version 2.0.0 [ZA6712_IE.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD)). (2017l). *Programme for the International Assessment of Adult Competencies (PIAAC), Italy log file*. Data file version 2.0.0 [ZA6712_IT.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017m). *Programme for the International Assessment of Adult Competencies (PIAAC), South Korea log file*. Data file version 2.0.0 [ZA6712_KR.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD)). (2017n). P*rogramme for the International Assessment of Adult Competencies (PIAAC), Netherlands log file*. Data file version 2.0.0 [ZA6712_NL.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017o). *Programme for the International Assessment of Adult Competencies (PIAAC), Norway log file.* Data file version 2.0.0 [ZA6712_NO.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017p). *Programme for the International Assessment of Adult Competencies (PIAAC), Poland log file.* Data file version 2.0.0 [ZA6712_PL.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017q). *Programme for the International Assessment of Adult Competencies (PIAAC), Slovakia log file*. Data file version 2.0.0 [ZA6712_SK.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

Organisation for Economic Co-operation and Development (OECD). (2017r). *Programme for the International Assessment of Adult Competencies (PIAAC), United States log file*. Data file version 2.0.0 [ZA6712_US.data.zip]. Cologne: GESIS Data Archive. doi:https://doi.org/10.4232/1.12955

PIAAC Expert Group in Problem Solving in Technology-Rich Environments. (2009). *PIAAC problem solving in technology-rich environments: A conceptual framework*. Paris: OECD Publishing.

Pohl, S., Ulitzsch, E., & von Davier, M. (2019). Using response times to model not-reached items due to time limits. *Psychometrika*, 1–29. https://doi.org/10.1007/s11336-019-09669-2.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rölke, H. (2012). *The ItemBuilder: A graphical authoring system for complex item development.* Paper presented at the World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, Montréal, Quebec, Canada.

Stelter, A., Goldhammer, F., Naumann, J., & Rölke, H. (2015). Die Automatisierung prozeduralen Wissens: Eine Analyse basierend auf Prozessdaten [The automation of procedural knowledge: An analysis based on process data]. In J. Stiller & C. Laschke (Eds.), *Berlin-Brandenburger Beitrage zur Bildungsforschung 2015: Herausforderungen, Befunde und Perspektiven Interdisziplinärer Bildungsforschung* (pp. 111–131). https://doi.org/10.3726/978-3-653-04961-9.

Tóth, K., Rölke, H., Goldhammer, F., & Barkow, I. (2017). Educational process mining: New possibilities for understanding students' problem-solving skills. In B. Csapó & J. Funke (Eds.), *The nature of problem solving: Using research to inspire 21st century learning* (pp. 193–209). Frankfurt am Main: Peter Lang. https://doi.org/10.1787/9789264273955-14-en.

Van der Linden, W. J. (2005). *Linear models for optimal test design*. New York: Springer.

Van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72*, 287–308. https://doi.org/10.1007/s11336-006-1478-z.

Van der Linden, W. J. (2008). Using response times for item selection in adaptive testing. *Journal of Educational and Behavioral Statistics, 33*, 5–20. https://doi.org/10.3102/1076998607302626.

Van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika, 73*, 365–384. https://doi.org/10.1007/s11336-007-9046-8.

Vörös, Z., & Rouet, J.-F. (2016). Laypersons' digital problem solving: Relationships between strategy and performance in a large-scale international survey. *Computers in Human Behavior, 64*, 108–116. https://doi.org/10.1016/j.chb.2016.06.018.

Weeks, J. P., von Davier, M., & Yamamoto, K. (2016). Using response time data to inform the coding of omitted responses. *Psychological Test and Assessment Modeling, 58*, 671–701.

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement, 43*, 19–38. https://doi.org/10.1111/j.1745-3984.2006.00002.x

Wise, S. L., & Gao, L. (2017). A general approach to measuring test-taking effort on computer-based tests. *Applied Measurement in Education, 30*, 343–354. https://doi.org/10.1080/08957347.2017.1353992.

Wise, S. L., Kuhfeld, M. R., & Soland, J. (2019). The effects of effort monitoring with proctor notification on test-taking engagement, test performance, and validity. *Applied Measurement in Education, 32*, 183–192. https://doi.org/10.1080/08957347.2019.1577248.

Yamamoto, K., & Lennon, M. L. (2018). Understanding and detecting data fabrication in large-scale assessments. *Quality Assurance in Education, 26*, 196–212. https://doi.org/10.1108/QAE-07-2017-0038.