Interdisciplinary frontiers: computer-based process data analysis in educational measurement

Björn Nicolay[a], Florian Krieger[b], and Samuel Greiff[a], [a] University of Luxembourg, School of Humanities, Education and Social Sciences, Department of Behavioral and Cognitive Sciences, Campus Belval, Maison des Sciences Humaines, 11, Porte des Sciences, Esch-sur-Alzette, Luxembourg; and [b] Technical University Dortmund, Department of Rehabilitation Sciences, Dortmund, Germany

In the overarching realm of educational measurement, the collection and analysis of data in order to gather knowledge about students' skill sets and their learning progress has always been of focal interest for educational stakeholders (OECD, 2013, 2017). This is particularly the case since such data have several implications that can shape students' future educational trajectories. For instance, a student's grade point average (GPA) to a large extent determines whether they are allowed to advance in their educational journey or are required to repeat a class or even an entire grade (e.g., Huebener and Marcus, 2017).

However, teachers, parents, and students themselves are not only interested in data indicating a passing or failing score in a particular subject or the GPA. Instead, data on factors that contribute to one's chances of meeting—or falling short of meeting—the standards required to achieve educational progress carry comparatively more fine-grained and thus crucially valuable information. Such data may include the time spent on individual learning exercises or the number of interactions made when working on a computerized training program focusing on the acquisition or refinement a specific skill (i.e., computer-based process data).

Such computer-based process data include all information collected during a student's interaction with a given computer program, ranging from time spent and number of clicks made on a particular page to the number and content of questions asked to a virtual tutor in a computerized learning simulation (Goldhammer et al., 2017). We will now provide an overview of the contents of this article, which, in brief, centers on the characteristics, history, contemporary use of process data analysis in educational measurement, a concrete example, and future challenges and directions in order to advance educational policy and practice.

Facets of computer-based process data analysis in educational measurement featured in this article

In this article, we first address the question of what computer-based process data are and illustrate their relevance for the educational measurement context (Section **The role of computer-based process data in educational measurement**). Thereafter, in **History of computer-based process data analysis in educational measurement** section, we present the historical evolution of computer-based process data analysis in educational measurement from its inception to its widespread use today. Subsequently, the availability of public computer-based process data repositories and the importance of computer-based process data in large-scale educational assessments will be described (Section **Public computer-based process data repositories**). In **The current state of computer-based process data analysis** section, we discuss the current state of computer-based process data analysis in educational measurement in greater depth with regard to three central facets: (4a) psychometric models; (4b) formative assessment; and (4c) summative assessment. Psychometric models are included as they serve as the statistical backbone for the thorough and valuable analysis and interpretation of computer-based process data in the educational context (particularly with regard to valid and reliable student assessment; DiCerbo et al., 2017; Mislevy et al., 2012). Meanwhile, formative and summative assessment represent today's two most widely used assessment techniques for monitoring students' learning progress, and have at the same time become a hallmark

International Encyclopedia of Education, Fourth Edition, 2023, 417–429

of evaluating and advancing educational practice (Gallardo, 2021; Harlen, 2011). Thereafter, in **Applied computer-based process data analysis in educational measurement: the example of complex problem solving** section, we provide an in-depth showcase of applied computer-based process data analysis in educational measurement by referring to the skill complex problem solving, which has been shown to be a key competence for students' educational success and beyond (Mainert et al., 2015; Schweizer et al., 2013). The article concludes with a summary of the relevant aspects of educational computer-based process data analysis and some final thoughts on current challenges and potential future directions for computer-based process data analysis in the educational measurement context (Section **Summary, challenges, and future directions of computer-based process data analysis in educational measurement**).

Notably, this article aims to present a thorough overview and theoretical background of process data analysis in educational measurement, and to provide indications for interested researchers and students on exemplary deeply vetted resources facilitating individual process data analysis endeavors. However, due to the vast amount of theory, practical aspects, and considerations involved in computer-based process data analysis, it is unrealistic to assume that this article provides information on each and every nuance of process data analysis in educational measurement. Therefore, we will at times point the interested reader to further in-depth scholarly sources, such as journal articles or books, about specific phenomena in educational process data analysis. Where applicable, these suggestions are given at the end of each respective section.

The role of computer-based process data in educational measurement

Computer-based process data are automatically stored in computerized log files (Zoanetti and Griffin, 2017) that contain raw information on each action a student performs with a software (e.g., a task or an exercise; see **Fig. 1**). Compared to other well-established types of process data in educational measurement research, such as think-aloud protocols or eye tracking (Branch, 2001; Horstmann et al., 2009), computer-generated process data are generally collected in a less "invasive" way while a student performs different actions on a computer; that means, the student is typically not explicitly aware of being monitored (Adams et al., 2015). In this regard, the generation and collection of such computer-based process data in digital gamified learning environments (i.e., educational game software) has also been termed "stealth assessment" (Shute, 2011; Shute et al., 2016b). The ability to automatically log computer-based process data in such a non-interfering way is a particularly valuable asset, since students' subjective reports of their perceived skill levels can be susceptible to bias and inaccuracy, as they put additional demands on the student and could interfere with actual processing of the task (e.g., Brown et al., 2015).

The thorough scrutiny of the vast amount of computer-based process data inherent in such log files (i.e., computer-based process data analysis) is typically referred to as data mining, which comprises various steps in order to arrive at data that is ready to be interpreted (Aldowah et al., 2019; Ali, 2013). As this process is generally complex and requires advanced methods, time and computer programming skills, it is usually performed by (semi-) automated programs provided by experts in the field (Gobert et al., 2013; Romero and Ventura, 2007, 2020; Slater et al., 2017). Importantly, these programs are tailored to produce the pieces of information desired for a given purpose, such as universities looking into reasons why students drop out of a given study program or teachers trying to uncover the reasons for their students' poor test results (Hsia et al., 2008; Mostow et al., 2005).


```
−<logEntry xsi:type="cbaloggingmodel:LogEntryTimeStamp" timeStamp="2016-03-21T08:58:56.302+0100">
  −<logEntry xsi:type="cbaloggingmodel:MicroDynButtonPressLogEntry" button="Execute" phase="control">
      <variable name="Pelikortit" userDefinedId="EndoA" value="22"/>
      <variable name="Pisteet" userDefinedId="EndoC" value="22"/>
      <variable name="Pelinappulat" userDefinedId="EndoB" value="22"/>
      <variable name="Vihreät pelimerkit" userDefinedId="ExoB" value="1"/>
      <variable name="Siniset pelimerkit" userDefinedId="ExoA" value="1"/>
      <variable name="Punaiset pelimerkit" userDefinedId="ExoC" value="1"/>
    </logEntry>
  </logEntry>
−<logEntry xsi:type="cbaloggingmodel:LogEntryTimeStamp" timeStamp="2016-03-21T08:58:57.128+0100">
  −<logEntry xsi:type="cbaloggingmodel:MicroDynButtonPressLogEntry" button="Execute" phase="control">
      <variable name="Pelikortit" userDefinedId="EndoA" value="24"/>
      <variable name="Pisteet" userDefinedId="EndoC" value="24"/>
      <variable name="Pelinappulat" userDefinedId="EndoB" value="24"/>
      <variable name="Vihreät pelimerkit" userDefinedId="ExoB" value="1"/>
      <variable name="Siniset pelimerkit" userDefinedId="ExoA" value="1"/>
      <variable name="Punaiset pelimerkit" userDefinedId="ExoC" value="1"/>
```


Fig. 1 Part of a raw log file in .xml format containing process data depicting the actions applied by a student during a complex problem solving task (see also **Summary, challenges, and future directions of computer-based process data analysis in educational measurement** section and **Fig. 2** of this article) in a computer-based large-scale assessment in Finland. Based on Vainikainen (2014).

History of computer-based process data analysis in educational measurement

The computerized administration of not only tests but also classroom learning exercises has become increasingly common since the late stages of the 20th century (Bunderson et al., 1989). This trend continues to be observed today (Danniels et al., 2020). Accordingly, the mid-1990s mark the inception of the first research articles referring to the use of computer-based process data to evaluate and improve policies in educational institutions (for an overview on data mining in the educational context between 1995 and 2005 see Romero and Ventura, 2007). One of the first articles explicitly using the term "computer-based process data," albeit in an organizational context, was published by Langley (1999). Soon thereafter, further researchers began scrutinizing computer-based process data from different types of log files for educational purposes (Srivastava et al., 2000), leading to rapid growth of the field. One of the earliest published computer-based process data analyses presents algorithms used to thoroughly map a variety of student characteristics (e.g., courses taken and credits earned in different disciplines) in order to more efficiently allocate staff and resources in educational institutions (Luan, 2002). Other pioneering studies address the leveraging of computer-based process data from digital learning and communication platforms for uploading presentation slides and exercises, submitting reports, and directly asking questions to peers and teachers (Mazza and Milani, 2005; Mostow, 2004; Paulsen, 2003).

The aforementioned selection of different ways to investigate and use computer-based process data for educational purposes further expanded over the first decade of the 21st century. During this period, studies began to focus on particular benefits for both students and educators, such as improving communication and adapting curricula by clustering and classifying students into different groups, for example (Romero and Ventura, 2007). A comprehensive overview of the most influential papers in the early years of computer-based process data research can be found in Baker and Yacef (2009).

At around the same time, the educational data mining sub-field known as process mining emerged, which focuses on the analysis computer-based process data (i.e., log files) with the aim of improving existing processes, or skills, in a variety of domains, from workers' deadline management in organizations to how people book vacation lodging online (Van der Aalst et al., 2010). While initially primarily targeted at business process models (Van der Aalst, 2012), process mining was rapidly extended to the educational domain. Existing research in this vein draws upon computer-based process data collected from students watching different educational video sequences, or from their interactions with digital learning platforms, in order to adapt said videos and platforms accordingly (Bogarín et al., 2014; Van der Aalst et al., 2015). Similar research has been conducted to increase the appeal of computer-based training programs based on inferences about people's motivation to take part in such programs drawn from computer-based process data (Cairns et al., 2015). Moreover, contemporary research investigating computer-based process data has targeted the evaluation and improvement of a variety of additional educational measurement facets. As such, recent research articles examine interactive learning environments, different assessment approaches, and individual students' skills that have been shown to be particularly related to educational success, and not seldom make use of publicly available computer-based process data repositories for their analyses (Cerezo et al., 2020; Greiff et al., 2015; Nicolay et al., 2021).

In sum, it can be seen that computer-based process data analysis for educational purposes has become more "fine-grained" in the last 15–20 years. In fact, an increased specialization of individual process data analyses has primarily been made possible due to technical advancements allowing researchers to capture the desired processes accurately using increasingly powerful computers and software (Adjerid and Kelley, 2018). Additionally, the availability of public process data repositories has put increased amounts of more easily accessible data at the disposal of both researchers and educational stakeholders.

Public computer-based process data repositories

Nearly simultaneously with the widespread recognition of the scholarly potential of computer-based process data analysis, researchers became particularly interested in using such data to improve educational practice (e.g., Branch, 2001). Initially, scientists engaging in computer-based process data analysis largely had to work with their own datasets. However, around 15 years ago, the first public computer-based process data repository was established, creating a new possibility for researchers to carry out computer-based process data analysis without being forced to collect their own data, and easing the comparison and verification of published results (Koedinger et al., 2008). At the same time, research began to increasingly rely on data from digital learning environments used for communication among teachers and students (Baker and Yacef, 2009).

Currently, several public computer-based process data repositories from different sources exist. We want to showcase two contexts of data sources, which are frequently employed by researchers.

First, numerous public repositories for computer-based process data exists from *online tutoring systems* (e.g., "ASSISTments," "DataShop," "vMS-Share," "KEEL"). These repositories are continuously growing, are freely accessible to researchers all over the world, and provide data from people's interactions with computer-based systems across multiple domains, either in their raw log file format, or in item-level responses per individual participant. Given the amount and variety of available data, these public repositories represent a crucial resource for the advancement of educational policy and practice (Alcalá-Fdez et al., 2011; Blonder et al., 2019; Feng et al., 2009; Koedinger et al., 2010).

Second, data from international large scales assessments (ILSAs) represent another valuable source for data mining. The two presumably most widely known ILSAs "Program for International Student Assessment" (PISA) and "Program for the International Assessment of Adult Competencies" (PIAAC) offer great insights into (educationally) relevant competencies, including reading ability, numeracy, or innovative competencies, like problem solving, among both school students (PISA) and adults (PIAAC) across the globe. Some of these competencies are assessed with computer-based assessments and thus, are stored in computer-based process data (Jude, 2016; Kirsch and Lennon, 2017). As the logged process data of both assessments are publicly available on the official website of the Organization for Economic Co-operation and Development (OECD; see https://www.oecd.org/pisa/data/ for PISA, and https://www.oecd.org/skills/piaac/data/piaaclogfiles/ for PIAAC), it is not surprising that multiple studies have used these data for fine-grained analyses of students' and adults' educational achievement (Goldhammer et al., 2020; Greiff et al., 2015; He et al., 2017; Liao et al., 2019; Teig et al., 2020). For both assessments, the process data are both available in raw log file format, and in item-based responses per country. Thus, it can be seen that the computer-based process data stored in the publicly available data repositories for PISA and PIAAC represent a valuable resource for scholars to learn more about how both students and adults interact with digital environments as well as their skills in key educational competencies.

The current state of computer-based process data analysis

Having provided a thorough overview of the purpose, history, and application of computer-based process data analysis in educational measurement, and introducing some of the most relevant publicly available computer-based process data repositories, we will now discuss the current state of computer-based process data analysis. In this regard, we will focus on three major facets of computer-based process data analysis due to their importance for successfully conducting and drawing valid inferences from computer-based process data analysis: (a) psychometric models (see 4.1), (b) formative assessment (see 4.2), and (c) summative assessment (see 4.3). In brief, psychometric models serve as statistical foundations for computer-based process data analysis. Formative assessment refers to the evaluation of students' learning progress during a particular learning exercise or course, while summative assessment measures students' learning progress after the learning exercise or course has been completed.

Psychometric models

Psychometrics can be defined as "the science of psychological assessment" (Rust and Golombok, 2014, p. 4) and is concerned with how psychology research measures designated phenomena to arrive at meaningful conclusions (Jones and Thissen, 2006). Computer-based process data pose several challenges for psychometric modeling that should be highlighted before moving on to psychometric models that have been specifically developed to deal with these challenges (Mislevy et al., 2012).

Firstly, computer-based process data differ markedly from "traditional" data (i.e., data that were not collected automatically via a digital entity during an assessment process). Such "traditional" data are usually limited to an absolute preference or performance score obtained from each individual participant over a given number of items. For instance, when analyzing such "traditional" data, it is nearly impossible to detect a learning effect for a participant over the course of multiple assessment items apart from a potentially higher overall success rate. Computer-based process data, on the other hand, might detect such a learning effect not only in the form of higher overall performance scores, but also using other indicators like consistently lower response times or fewer intermediate errors committed over several items.

Secondly, logged process data are not limited to a particular type of data such as responses on a Likert scale or number of items correctly recalled from memory. Instead, a single log file contains numerous data points in various formats, such as written information the participant typed, time stamps indicating when the participant started and completed a task, and the exact times and instances the participant clicked on each button shown on the computer screen. Therefore, compared to traditional data, computer-based process data pose unique challenges for psychometric models due to their fine-grained nature.

Today, however, a variety of different psychometric models (i.e., statistical frameworks aiming to explain how data are generated and can subsequently be analyzed in a valid and reliable manner; American Psychological Association, 2020) suitable for computer-based process data analysis are available that seek to address these challenges. One example is the evidence-centered design (ECD) framework, which has received increasing attention for computer-based process data analysis in the educational context (Mislevy et al., 2003). The ECD framework encompasses several different parts. The *student model* consists of the variables to be measured, while the *task model* refers to the situations in which these variables can be observed. Notably, the psychometric model, together with evidence identification rules, serves as a link between the two former models, creating a final evidence model. Based on this ECD framework, Levy (2020) identifies the four psychometric categories of preparing data, analyzing data, presenting data, and situating the work in each of these three categories. In order to account for the varying nature of computer-based process data, Levy (2020) argues that the interdependency between such process data (indicative of a person's behavior during the task completion process) and product data (an end result or performance score after task completion) needs to be considered. This can be accomplished, for instance, by adding variables from the student model in order to distinguish factors related to the process from those constituting the product, and additional related techniques (p. 223).

Next to the ECD, the Extended Learning and Assessment System (XLAS) framework (Von Davier et al., 2019) represents a second psychometric model for computer-based process data. The developers of this framework argue that the three dimensions of *assessment*, *learning*, and *navigation* must be considered when working with multimodal data such as computer-based process data. They additionally emphasize that, when trying to recover information from computer-based process data, how to extract and create meaningful results from initial raw and unorganized bunches of data should be prioritized.

In addition, Kroehne and Goldhammer (2018) propose using finite state machines to extract the desired data points from initially cumbersome-to-analyze log files. Such finite state machines act as a measurement tool of how the participant interacts with the computer during a specific part of the overall assessment; thus, they act as encoders of the information stored in a given log file. This specific part (e.g., one item within a large questionnaire) represents the finite state, which can be defined theoretically a priori, but also changed later on—for example, to achieve comparability with other assessment approaches incorporating similar action sequences. Importantly, participants' behavior as they progress through this collection of subsequent finite items can then be investigated not only on the within-person level, but also across the different items. Kroehne and Goldhammer (2018) illustrate this procedure using computer-based process data from PISA, discovering that their finite state machine approach produces valid results.

Taken together, although the complexity of computer-based process data poses a unique challenge for researchers, several psychometric models are currently available to facilitate scholarly work with such data. Indeed, psychometric model approaches such as XLAS, ECD, and finite state machines represent valuable tools for both educational stakeholders and scholars. Importantly, as psychometric models are generally based on a theoretical foundation, they are only as useful as their underlying theory. Thus, a strong and well-established theory serves as the best precursor for the development of a meaningful and useable psychometric model (Rupp et al., 2012; Zehner et al., 2020). Consequently, building one's psychometric model on a sound theoretical foundation crucially facilitates formative and summative assessment of students based on computer-based process data (Drachsler and Goldhammer, 2020; Zhu et al., 2016).

Formative assessment

Evaluating students' learning progress by means of different assessment techniques has always been an integral part of educational measurement. Nevertheless, formative assessment has received increasing attention in education since the beginning of the 21st century (e.g., Bell and Cowie, 2001). Formative assessment entails the evaluation of students' learning in order to adapt learning policies (e.g., curricula) both on an individual level as well as on a broader class or school level (Van der Kleij et al., 2015). Formative assessment practices are widely known as "assessment for learning" and typically performed during the course of a learning activity or semester rather than afterward, as is the case for summative assessment (Saeed et al., 2018; see 4.3). In addition, a core purpose of formative assessment is to support students' learning by giving and receiving feedback. This feedback does not inherently center on monitoring a student's progress with respect to the learning content, but rather seeks to provide suggestions for improvement with regard to a student's learning strategies (Gikandi et al., 2011). Furthermore, formative assessment can also be computer-based (Bergeson, 2019; Wijesooriya et al., 2015). As such, it can be put into practice by asking students to carry out a classroom task in several smaller groups, for example, and subsequently having the students provide feedback to their peers in the same group. This not only allows students to point out aspects of tasks their peers are approaching differently, but also directly engage in a discussion about how these different learning strategies might be more or less likely to (in)efficiently lead to (in)correct results. Moreover, such discussions can increase students' knowledge about learning strategies and their flexibility in applying and adapting them. Overall, formative assessment represents an important complement to summative assessment, as it can significantly help students achieve higher performance scores in summative assessment tests (Harlen, 2011; Zheng et al., 2019).

The analysis of computer-based process data from formative assessment contexts is spurred by its increasing relevance in educational contexts. Several studies have developed digital learning environments that provide feedback to students throughout their interaction with the environment, while simultaneously yielding a rich set of computer-based process data. For instance, analyzing computer-based process data from the personalized e-learning system PELS, which specifically tailors the content and frequency of feedback to a student's individual needs, revealed a significant increase in students' mathematical abilities after completing learning sessions in the digital learning environment involving more immediate feedback (Chen and Chen, 2009). Similar results were obtained for a related program in the field of mathematics called MATHia, which students used over the course of an entire school year (Zheng et al., 2019). In their study, Zheng et al. (2019) analyzed the computer-based process data employing different statistical models, which produced different results. Thus, the authors emphasize the importance of specifying the statistical methodology of choice in a given context prior to diving into a fine-grained investigation of computer-based process data (e.g., cluster analysis versus hierarchical modeling; Miller et al., 2015). Likewise, researchers should carefully consider how to use and interpret student scores obtained from computer-based process data in order to ensure the validity of formative assessments and their results (Hopster-den Otter et al., 2019).

Alongside established school subjects such as mathematics, domain-general 21st century skills (i.e., skills such as complex problem solving that have been identified as especially relevant for today's learners; Trilling and Fadel, 2009) have been located at the core of digital learning environments producing computer-based process data. In this regard, an analytical framework has been established addressing how to immediate provide feedback to students while they are still working on a given task (Choi and Cho, 2020). This allows students' learning paths to be modeled in more depth, tracing not only the final result of a particular task, but also the individual steps the student took to arrive at this final result. Moreover, Choi and Cho (2020) analyzed the computer-based process data from their formative assessment methodology by contrasting the observed learning trajectory of individual students to a predefined expected learning trajectory at several time points during a task. This allowed the specific time points at which feedback should be provided to students to be identified, leading to improved feedback timing and content (Choi and Cho, 2020).

Summative assessment

At present, the most widely applied assessment technique in Western classrooms is summative assessment (Saeed et al., 2018). Summative assessment refers to assessing students' learning outcomes once at the end of a course in order to make decisions about the students' ability to transition to other courses or the next grade level (Anthony and Susan, 2005). **Table 1** can help readers understand the differences between formative and summative assessment at a glance.

One example of summative assessment in practice is to require university students to complete a final exam in an introductory module at the end of the semester, and subsequently using each student's score on this final exam to determine whether they are permitted to enroll in the advanced module next semester. In general, summative assessments are deployed for purposes at the individual student level, such as admissions, ability grouping, grading, norming, or mastery of a particular subject (McGarr and Clifford, 2013; Van Groen and Eggen, 2020).

Multiple research studies have already used computer-based process data analysis to shed light on factors influencing students' summative assessment results. For instance, Aojula et al. (2006) compared the results of a computer-based summative assessment of undergraduate students to teachers' marking by hand and to the results obtained from multiple-choice marking machines that read students' answer sheets. They demonstrate that using computer-based process data analysis represents a viable and beneficial alternative to traditional summative assessment modalities, given its higher practicality and accuracy (e.g., the results are provided directly and automatically and do not need to be calculated by hand). Furthermore, Gocheva-Ilieva et al. (2020) employed common data mining techniques such as cluster analysis to evaluate the impact of different elements of the summative assessment approach on students' final grades in mathematics, discovering a significant influence of students' sex and selected elective subjects. Likewise, computer-based process data from a large-scale summative assessment in France were used to examine the potential benefit of students' employment of strategies such as trial and error versus a planned, structured approach when solving a computer-based mathematics problem (Salles et al., 2020). The authors concluded that although neither of the two strategic approaches resulted in better or worse overall performance, all students could validly be classified into one of the two strategy groups using the employed cluster analysis. Another recently published study presents the assessment program Winsight®, which targets potential improvements in individual summative assessment components (e.g., by providing information about students' proficiency and ensuring full accessibility for students with disabilities or limited knowledge of the language of instruction) in order to foster student learning (Stone and Wylie, 2019). The program primarily seeks to funnel such information to educational stakeholders such as school headmasters and curricular decision-makers. The authors apply their new approach using computer-based process data analysis in order to assess test accessibility (i.e., how different populations of students navigate through the test, select and refine their answers) and thus provide good validity indicators for their assessment program.

In sum, we have seen that caution is advised when analyzing computer-based process data to ensure that valid, meaningful and interpretable results are obtained that carry useful implications for education. However, when suitable statistical methodology and psychometric models capable of accurately capturing and presenting the desired information are applied, computer-based process data analysis represents a valuable tool for the advancement of both formative and summative assessment practices. This includes but is not limited to providing feedback about and evaluating students' learning trajectories.

Table 1 Overview of the purpose, administration, usability, and examples of formative and summative assessment in education.

	Formative assessment	*Summative assessment*
Goal	Providing feedback and facilitating learning and instruction	Assessing students' knowledge of a particular subject or domain
Time of application	During a learning exercise or course	After a learning exercise or course
Use for teachers	Evaluating students' learning methods and the quality of the learning material, adaptation of exercises and curricula	Understanding individual and class performance, comparing scores to other schools, grading
Use for students	Assessing one's own learning methods in relation to peers, self-evaluation	Evaluating one's own progress in relation to peers and curricular standards
Examples	Self-tests, practice exam	Final exam, individual report

Applied computer-based process data analysis in educational measurement: the example of complex problem solving

At this point, having presented a general overview of the current state of computer-based process data analysis in educational measurement, we would like to illustrate how computer-based process data analysis can be used in practice to fill knowledge gaps regarding how students apply a skill that has been shown to be particularly relevant for educational success and beyond: the domain-general 21st century skill complex problem solving (CPS).

CPS skills can be defined as the ability to solve problems that possess unique features, such as hidden relationships between variables, multiple goals to be reached simultaneously, and the potential for variable values to change dynamically at any given time during the solution process (Greiff et al., 2013a,b). One example of a complex problem could be trying to use a recently updated self-service ticket machine to purchase a train ticket, an example that has already been used in PISA (OECD, 2014; here CPS was labeled as creative problem solving). In order to ultimately make the correct purchase, it is crucial to apply a strategic approach (i.e., deliberately clicking on different buttons on the machine) to build up one's knowledge instead of performing random clicks (i.e., applying the vary-one-thing-at-a-time or VOTAT strategy, also referred to as the control of variables or CVS strategy; Schwichow et al., 2016).

Computer-based microworlds have become the current standard for assessing students' CPS skills (e.g., Stadler et al., 2015). These microworlds involve different hypothetical scenarios as complex problems that must be solved, such as creating a lemonade using different ingredients with arbitrary labels or evaluating the impact of changing the dosage of several unknown types of medication on different body functions (Greiff et al., 2013a,b). In CPS assessment, students can explore the initially hidden relationships among the variables within a given item by moving sliders and clicking different buttons. These actions, as well as others, can be used for process data analysis. See Fig. 2 below for an exemplary CPS item.

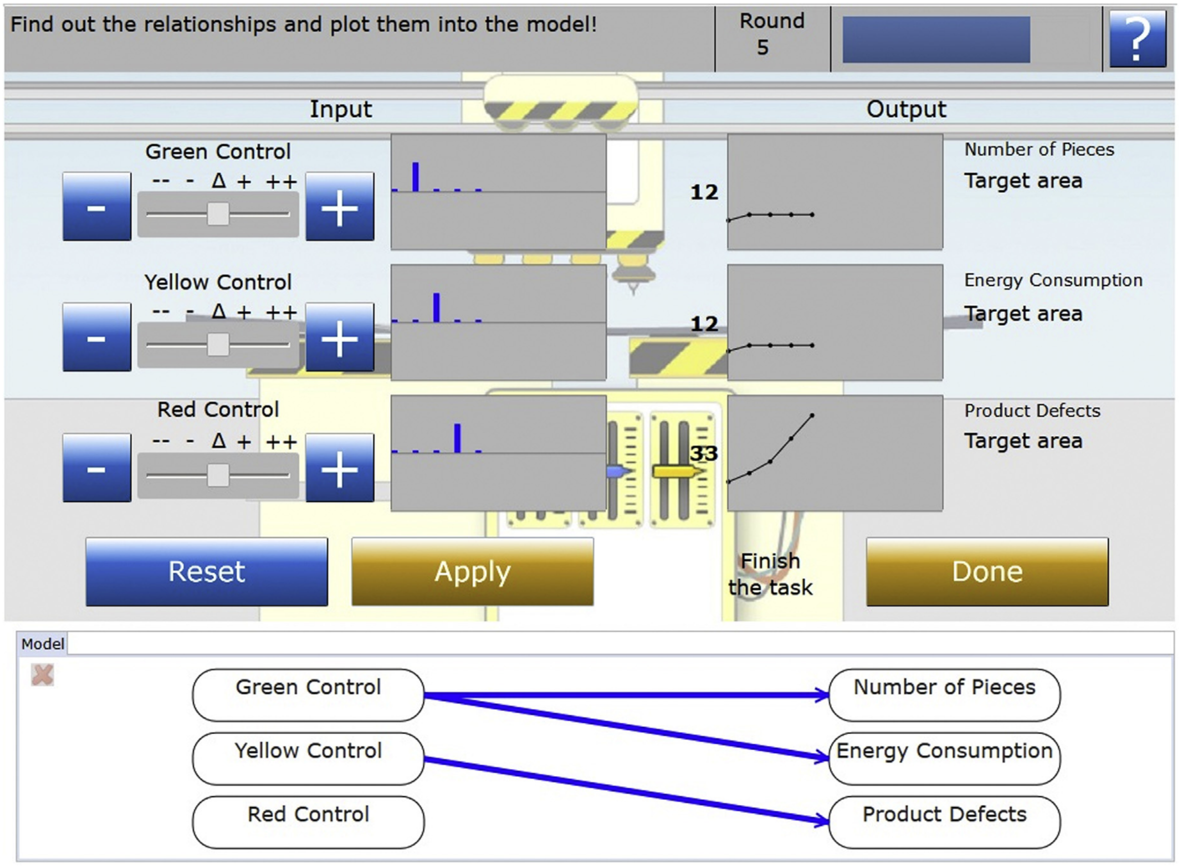

Fig. 2 A complex problem microworld involving multiple variables with hidden relationships (see also **Fig. 1** in **The role of computer-based process data in educational measurement** section for a corresponding log file containing process data). After the variable relationships have been discovered in the problem space (top part) by applying strategies such as VOTAT, the student is asked to indicate each present relationship in the model space (bottom part).

Importantly, computer-based process data are collected and stored in log files while students work on the complex problems within such a microworld. These log files capture all relevant components of a student's interaction with a given problem, including time spent on an item, all variable manipulations and their corresponding time stamps, and whether the student achieved or at least approximated the predefined goals (Xu et al., 2018). A common approach to extracting valuable information from this computer-based process data after CPS assessment is to write an automated program that iterates through all log files and parses the relevant data points for further subsequent statistical analyses. For instance, if data are stored in XML-format, this can be achieved using so-called XML parsers (i.e., programs that automatically extract the desired information from raw log files; e.g., Applen and McDaniel, 2009).

The importance of CPS for both educational and career success as the ability to manage novel and dynamic situations in a systematic way has been discussed in multiple studies (Schweizer et al., 2013; Wüstenberg et al., 2012). Thus, it does not come as a surprise that a large body of research on CPS using computer-based process data from large-scale assessments, including PISA, has accumulated in recent years (Greiff et al., 2015; Han et al., 2019; Liu et al., 2020). For instance, research has established the particular benefit of applying strategies such as the aforementioned VOTAT strategy alone or in concert with other related strategies for successful CPS performance (Molnár and Csapó, 2018). Additionally, the time students spend on individual complex problems has been shown to influence their probability of solving the problem successfully or unsuccessfully (Scherer et al., 2015). A study by Ren et al. (2019) uncovered how students can successfully balance multiple goals to be achieved simultaneously in CPS. Additional research has sought to identify different levels of CPS proficiency in students, for example based on how they approach a CPS assessment test (Greiff et al., 2015; Stadler et al., 2020). Eichmann et al. (2020) used PISA computer-based process data to uncover differences in students' CPS performance based on their sex and ethnic background: the higher CPS success rates found among boys compared to girls were attributable to differences in exploration behavior between the sexes, whereas the behavioral differences investigated were unable to account for the performance differences between students with versus without migration backgrounds.

In summary, as demonstrated by the selected research endeavors described above, computer-based process data analysis has greatly supported the investigation of students' competence levels in crucial skills such as CPS (for an overview of computer-based process data use in CPS, see also Herde et al., 2016). At the same time, computer-based process data analysis has helped to uncover opportunities to improve students' skills in these areas by means of training programs in digital learning environments (e.g., Azevedo, 2007). Moreover, as computers and digital learning and assessment tools become increasingly prevalent in the educational context, we can expect further advancements based on computer-based process data analysis in the near future. Consequently, we will now discuss potential challenges and future focus areas for computer-based process data analysis, after providing a broad summary of its current state of application in the educational measurement domain.

Summary, challenges, and future directions of computer-based process data analysis in educational measurement

We have seen that computer-based process data analysis in educational measurement carries a variety of benefits and challenges simultaneously for both educational stakeholders and researchers. Despite the fact that this promising research tool emerged not long ago, it has already produced several noteworthy advancements, as described throughout this article. Computer-based process data analysis has evolved into a versatile method for evaluating students' learning progress in order to create specifically targeted assessment and training programs for educationally relevant competencies. As such, computer-based process data analysis represents an important means of drawing inferences about students' performance in both formative and summative assessment contexts (Greiff, 2020). On the other hand, the increasingly fine-grained, rich nature of computer-based process data necessitates the development and continuous refinement of psychometric models that are able to extract meaningful information from increasingly complex raw log files. Today, computer-based process data analysis has become an established and crucial tool for improving educational measurement across the globe. Moreover, as technology continues to advance, we can expect that use of computer-based process data analysis in and beyond the educational context will expand rather than contract in upcoming years (Bergner and von Davier, 2019; Dosta et al., 2020).

Turning to potential future directions for computer-based process data analysis in educational measurement, three aspects must be considered: (a) the ongoing changing nature of assessment; (b) the evolution of digital learning environments; and (c) the increasing richness and diversity of computer-based process data, which calls for powerful and flexible statistical analysis tools. The opportunities and challenges associated with each of these aspects will be considered individually below, in order to provide some suggestions about where the use of computer-based process data analysis for educational purposes may be headed in the near future.

Firstly, we have already discussed the increasing use of computer-based process data analysis in both formative and summative assessment contexts. Both approaches represent particularly relevant educational measurement practices, and, today, they are frequently used in conjunction (Owen et al., 2014; Yang and Li, 2018). In addition, diagnostic and portfolio assessment are also becoming increasingly relevant (Chen et al., 2007; Saeed et al., 2018). Diagnostic assessment is strongly related to formative assessment, and primarily seeks to unravel how well students are prepared for the specific difficulties they experience when working on a particular task (Csapó and Molnár, 2019). In portfolio assessment, students hand in different assignments to their teacher and are then given the possibility to revise these assignments based on their teacher's feedback. After these revisions have been made, each assignment is then handed in again and added to the student's portfolio, which serves as the primary method of assessing their overall learning progress over the course of a full semester or school year. As such, portfolio assessment features elements of both formative and summative assessment (Steen-Utheim and Hopfenbeck, 2019). Thus, as educational practice moves away from exclusive reliance on summative assessment, additional assessment approaches have emerged in recent years and are likely to continue to emerge and be refined in the future. These developments will involve the collection of a multitude of computer-based process data, and new types of computer-based process data will become available to both educational stakeholders and researchers (Shute et al., 2016a).

Secondly, throughout this article, we have presented several digital learning environments that have been developed to guide students through the acquisition and/or refinement of a specific skill. Over the last decade, these digital learning environments have largely evolved into digital game-based learning environments (Loh, 2012). The game-based nature of these environments is beneficial due to students' high levels of satisfaction when working with such programs, thereby keeping students interested, keen to learn, and entertained at the same time (Sawyer et al., 2018). Such digital game-based learning environments inherently produce a considerable amount of computer-based process data. For instance, in a meta-analysis of game-based learning, Qian and Clark (2016) discovered that asking students to work with digital game-based learning environments led to improved levels of critical thinking, creativity, communication, and collaboration. We can therefore see these types of digital learning environments being used increasingly often in educational contexts in upcoming years.

All in all, as with the observed changes in assessment approaches described above, the fact that digital learning environments are continuously evolving will impose new challenges but also new opportunities for computer-based process data analysis in educational measurement. These noteworthy developments in assessment type and methodology will lead to the availability of an increased amount of even more fine-grained computer-based process data in the future. This brings us to the final, and most crucial, aspect to be considered: the trade-off between the richness of computer-based process data and the power and suitability of statistical tools to analyze these data in a meaningful way. Indeed, as a result of the increasingly complex and comprehensive assessment approaches employed in educational contexts, the log files representing the basis for computer-based process data analysis are becoming increasingly fine-grained and difficult to scrutinize (Kumar and Chadha, 2011). Moreover, how to address the potential issue of missing and/or inaccurate data has been discussed in multiple studies, highlighting the importance of quality assurance in computer-based process data analysis (Weeks et al., 2016; Yamamoto and Lennon, 2018). Furthermore, in addition to computer-based process data generated by students completing tasks on laptops and tablets, which can at this point be seen as rather "traditional," other avenues providing complementary computer-based process data are likely to evolve in the future. For instance, technology-based process data from eye tracking (Bergner and von Davier, 2019; Emerson et al., 2020) and physiological sources such as blood pressure, heart rate, and electrodermal activity may be collected as a student works on a digital learning environment task (Sharma et al., 2020; Vrzakova et al., 2020). Moreover, computer-based process data from outside school, including students' interactions with smartphone learning apps they engage with in their leisure time, present an additional source of information relevant for their educational progress (Markowetz et al., 2014; Sin and Muthu, 2015). In other words, the current and potential future sources of computer-based process data for educational purposes seem almost unlimited. Consequently, it will become paramount for educational stakeholders to prioritize the analysis of certain kinds of computer-based process data over others due to the increasing variety of available information. In turn, the creation and refinement of statistical tools for successful, valid, and interpretable computer-based process data analysis will be necessary. As discussed in this article, several such tools already exist today, and we are confident that their development and refinement will continue in the future.

Overall, in its current form, computer-based process data analysis represents a valuable tool for both educational stakeholders and researchers to evaluate and advance educational measurement policies and practices. Since its inception in the final stages of the 20th century, computer-based process data analysis has become a standard method for investigating students' learning progress and learning success. Given the promising possibilities of this evolving technology, we expect computer-based process data analysis to become increasingly present in the future in the everyday lives of educational decision-makers, teachers, and researchers, who share the common goal of providing today's students with a state-of-the-art education in order to support and prepare them for the challenges they will face throughout and after their educational journey.

Acknowledgments

This research was supported by a grant from the Fonds National de la Recherche Luxembourg (Luxembourgish National Research Found) [CORE "TRIOPS"].

References

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., Care, E., 2015. Automatic coding procedures for collaborative problem solving. In: Assessment and Teaching of 21st Century Skills. Springer, Dordrecht, pp. 115–132.

Adjerid, I., Kelley, K., 2018. Big data in psychology: a framework for research advancement. Am. Psychol. 73 (7), 899–917. https://doi.org/10.1037/amp0000190.

Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F., 2011. Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. J. Mult. Valued Log. Soft Comput. 17.

Aldowah, H., Al-Samarraie, H., Fauzy, W.M., 2019. Educational data mining and learning analytics for 21st century higher education: a review and synthesis. Telematics Inf. 37, 13–49. https://doi.org/10.1016/j.tele.2019.01.007.

Ali, M.M., 2013. Role of data mining in education sector. Int. J. Comput. Sci. Mobile Comput. 2 (4), 374–383.

American Psychological Association, 2020. Psychometric Model. https://dictionary.apa.org/psychometric-model.

Anthony, J.N., Susan, M.B., 2005. Education Assessment of Students. Pearson Education Ltd.

Aojula, H., Barber, J., Cullen, R., Andrews, J., 2006. Computer-based, online summative assessment in undergraduate pharmacy teaching: the Manchester experience. Pharm. Educ. 6 (4), 229–236. https://doi.org/10.1080/15602210600886209.

Applen, J.D., McDaniel, R., 2009. The Rhetorical Nature of XML: Constructing Knowledge in Networked Environments. Routledge.

Azevedo, R., 2007. Understanding the complex nature of self-regulatory processes in learning with computer-based learning environments: an introduction. Metacognition Learn. 2 (2–3), 57–65. https://doi.org/10.1007/s11409-007-9018-5.

Baker, R.S., Yacef, K., 2009. The state of educational data mining in 2009: a review and future visions. J. Educ. Data Min. 1 (1), 3–17. https://doi.org/10.5281/zenodo.3554657.

Bell, B., Cowie, B., 2001. The characteristics of formative assessment in science education. Sci. Educ. 85 (5), 536–553. https://doi.org/10.1002/0sce.1022.

Bergeson, K.T., 2019. Reading specialists use verbal protocols as a formative assessment tool. Read. Teach. 73 (2), 185–193. https://doi.org/10.1002/trtr.1815.

Bergner, Y., von Davier, A.A., 2019. Process data in NAEP: past, present, and future. J. Educ. Behav. Stat. 44 (6), 706–732. https://doi.org/10.3102/1076998618784700.

Blonder, N., Orsburn, B.C., Blonder, J., Gonzalez, C.A., 2019. Visual mass-spec share (vMS-Share): a new public web-based mass spectrometry visualization and data mining repository. J. Proteonomics Bioinf. 12. https://doi.org/10.4172/0974-276X.1000495.

Bogarín, A., Romero, C., Cerezo, R., Sánchez-Santillán, M., 2014. Clustering for improving educational process mining. In: Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, pp. 11–15. https://doi.org/10.1145/2567574.2567604.

Branch, J.L., 2001. Junior high students and think alouds: generating information-seeking process data using concurrent verbal protocols. Libr. Inf. Sci. Res. 23 (2), 107–122. https://doi.org/10.1016/S0740-8188(01)00065-2.

Brown, G.T., Andrade, H.L., Chen, F., 2015. Accuracy in student self-assessment: directions and cautions for research. Assess Educ. Princ. Pol. Pract. 22 (4), 444–457. https://doi.org/10.1080/0969594X.2014.996523.

Bunderson, C.V., Inouye, D.K., Olsen, J.B., 1989. The four generations of computerized educational measurement. In: Linn, R.L. (Ed.), Educational Measurement, third ed. American Council on Education, pp. 367–407.

Cairns, A.H., Gueni, B., Fhima, M., Cairns, A., David, S., Khelifa, N., 2015. Process mining in the education domain. Int. J. Adv. Intell. Syst. 8 (1), 219–232.

Cerezo, R., Bogarín, A., Esteban, M., Romero, C., 2020. Process mining for self-regulated learning assessment in e-learning. J. Comput. High Educ. 32 (1), 74–88. https://doi.org/10.1007/s12528-019-09225-y.

Chen, C.M., Chen, M.C., 2009. Mobile formative assessment tool based on data mining techniques for supporting web-based learning. Comput. Educ. 52 (1), 256–273. https://doi.org/10.1016/j.compedu.2008.08.005.

Chen, C.M., Chen, M.C., Li, Y., 2007. Mining key formative assessment rules based on learner profiles for web-based learning systems. In: Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies, Niigata, 2007, pp. 584–588. https://doi.org/10.1109/ICALT.2007.189.

Choi, Y., Cho, Y.I., 2020. Learning analytics using social network analysis and Bayesian network analysis in sustainable computer-based formative assessment system. Sustainability 12 (19), 7950. https://doi.org/10.3390/su12197950.

Csapó, B., Molnár, G., 2019. Online diagnostic assessment in support of personalized teaching and learning: the eDia system. Front. Psychol. 10, 1522. https://doi.org/10.3389/fpsyg.2019.01522.

Danniels, E., Pyle, A., DeLuca, C., 2020. The role of technology in supporting classroom assessment in play-based kindergarten. Teach. Teach. Educ. 88, 102966. https://doi.org/10.1016/j.tate.2019.102966.

DiCerbo, K., Shute, V., Kim, Y.J., 2017. The future of assessment in technology rich environments: psychometric considerations. In: Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy. Springer International, pp. 1–21.

Dosta, M., Litster, J.D., Heinrich, S., 2020. Flowsheet simulation of solids processes: current status and future trends. Adv. Powder Technol. 31 (3), 947–953. https://doi.org/10.1016/j.apt.2019.12.015.

Drachsler, H., Goldhammer, F., 2020. Learning analytics and eAssessment—towards computational psychometrics by combining psychometrics with learning analytics. In: Burgos, D. (Ed.), Radical Solutions and Learning Analytics. Lecture Notes in Educational Technology. Springer, Singapore, pp. 67–80. https://doi.org/10.1007/978-981-15-4526-9_5.

Eichmann, B., Goldhammer, F., Greiff, S., Brandhuber, L., Naumann, J., 2020. Using process data to explain group differences in complex problem solving. J. Educ. Psychol. 112 (8), 1546–1562. https://doi.org/10.1037/edu0000446.

Emerson, A., Cloude, E.B., Azevedo, R., Lester, J., 2020. Multimodal learning analytics for game-based learning. Br. J. Educ. Technol. 51 (5), 1505–1526. https://doi.org/10.1111/bjet.12992.

Feng, M., Heffernan, N.T., Koedinger, K.R., 2009. Addressing the assessment challenge in an intelligent tutoring system that tutors as it assesses. J. User Model. User Adapt. Interact. 19, 243–266. https://doi.org/10.1007/s11257-009-9063-7.

Gallardo, K., 2021. The importance of assessment literacy: formative and summative assessment instruments and techniques. In: Babo, R., Dey, N., Ashour, A.S. (Eds.), Workgroups eAssessment: Planning, Implementing and Analysing Frameworks. Intelligent Systems Reference Library, vol. 199. Springer, pp. 3–25. https://doi.org/10.1007/978-981-15-9908-8_1.

Gikandi, J.W., Morrow, D., Davis, N.E., 2011. Online formative assessment in higher education: a review of the literature. Comput. Educ. 57 (4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004.

Gobert, J.D., Sao Pedro, M., Raziuddin, J., Baker, R.S., 2013. From log files to assessment metrics: measuring students' science inquiry skills using educational data mining. J. Learn. Sci. 22 (4), 521–563. https://doi.org/10.1080/10508406.2013.837391.

Gocheva-Ilieva, S., et al., 2020. Data mining for statistical evaluation of summative and competency-based assessments in mathematics. In: Martínez Álvarez, F., Troncoso Lora, A., Sáez Muñoz, J., Quintián, H., Corchado, E. (Eds.), International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019), Advances in Intelligent Systems and Computing, vol. 951. Springer, pp. 207–216. https://doi.org/10.1007/978-3-030-20005-3_21.

Goldhammer, F., Naumann, J., Rölke, H., Stelter, A., Tóth, K., 2017. Relating product data to process data from computer-based competency assessment. In: Competence Assessment in Education. Springer, Cham, pp. 407–425.

Goldhammer, F., Hahnel, C., Kroehne, U., 2020. Analyzing log file data from PIAAC. In: Large-Scale Cognitive Assessment. Springer, Cham, pp. 239–269.

Greiff, S., Holt, D.V., Funke, J., 2013a. Perspectives on problem solving in cognitive research and educational assessment: analytical, interactive, and collaborative problem solving. J. Probl. Solving 5, 71–91. https://doi.org/10.7771/1932-6246.1153.
Greiff, S., Wüstenberg, S., Holt, D.V., Goldhammer, F., Funke, J., 2013b. Computer-based assessment of complex problem solving: concept, implementation, and application. Educ. Technol. Res. Dev. 61 (3), 407–421. https://doi.org/10.1007/s11423-013-9301-x.
Greiff, S., Wüstenberg, S., Avvisati, F., 2015. Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. Comput. Educ. 91, 92–105. https://doi.org/10.1016/j.compedu.2015.10.018.
Greiff, S., 2020. Technology-based assessment in 21st century education. Hrsg. In: Zender, R., Ifenthaler, D., Leonhardt, T., Schumacher, C. (Eds.), DELFI 2020—Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V. Bonn: Gesellschaft für Informatik e.V., pp. 21–22.
Han, Z., He, Q., von Davier, M., 2019. Predictive feature generation and selection using process data from PISA interactive problem-solving items: an application of random forests. Front. Psychol. 10, 2461. https://doi.org/10.3389/fpsyg.2019.02461.
Harlen, W., 2011. On the relationship between assessment for formative and summative purposes. In: Gardner, J. (Ed.), Assessment and Learning. Sage Publishing, pp. 95–110. https://doi.org/10.4135/9781446250808.n6.
He, Q., von Davier, M., Greiff, S., Steinhauer, E.W., Borysewicz, P.B., 2017. Collaborative problem solving measures in the Programme for International Student Assessment (PISA). In: Innovative Assessment of Collaboration. Springer, Cham, pp. 95–111. https://doi.org/10.1007/978-3-319-33261-1_7.
Herde, C.N., Wüstenberg, S., Greiff, S., 2016. Assessment of complex problem solving: what we know and what we don't know. Appl. Meas. Educ. 29 (4), 265–277. https://doi.org/10.1080/08957347.2016.1209208.
Hopster-den Otter, D., Wools, S., Eggen, T.J., Veldkamp, B.P., 2019. A general framework for the validation of embedded formative assessment. J. Educ. Meas. 56 (4), 715–732. https://doi.org/10.1111/jedm.12234.
Horstmann, N., Ahlgrimm, A., Glöckner, A., 2009. How distinct are intuition and deliberation? An eye-tracking analysis of instruction-induced decision modes. Judgm. Decis. Mak. 4, 335–354.
Hsia, T.C., Shie, A.J., Chen, L.C., 2008. Course planning of extension education to meet market demand by using data mining techniques—an example of Chinkuo Technology University in Taiwan. Expert Syst. Appl. 34 (1), 596–602. https://doi.org/10.1016/j.eswa.2006.09.025.
Huebener, M., Marcus, J., 2017. Compressing instruction time into fewer years of schooling and the impact on student performance. Econ. Educ. Rev. 58, 1–14. https://doi.org/10.1016/j.econedurev.2017.03.003.
Jones, L.V., Thissen, D., 2006. 1 A history and overview of psychometrics. Handb. Stat. 26, 1–27. https://doi.org/10.1016/S0169-7161(06)26001-2.
Jude, N., 2016. The assessment of learning contexts in PISA. In: Kuger, S., Klieme, E., Jude, N., Kaplan, D. (Eds.), Assessing Contexts of Learning. Methodology of Educational Measurement and Assessment. Springer, Cham. https://doi.org/10.1007/978-3-319-45357-6_2.
Kirsch, I., Lennon, M.L., 2017. PIAAC: a new design for a new era. Large Scale Assess. Educ. 5 (1), 11. https://doi.org/10.1186/s40536-017-0046-6.
Koedinger, K.R., Cunningham, K., Leber, B., 2008. An open repository and analysis tools for fine-grained, longitudinal learner data. In: Proceedings of the 1st International Conference on Educational Data Mining, pp. 157–166.
Koedinger, K.R., Baker, R.S., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J., 2010. A data repository for the EDM community: the PSLC DataShop. In: Handbook of Educational Data Mining, vol. 43. CRC Press, pp. 43–56.
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.
Kumar, V., Chadha, A., 2011. An empirical study of the applications of data mining techniques in higher education. Int. J. Adv. Comput. Sci. Appl. 2 (3), 80–84.
Langley, A., 1999. Strategies for theorizing from process data. Acad. Manag. Rev. 24 (4), 691–710.
Levy, R., 2020. Implications of considering response process data for greater and lesser psychometrics. Educ. Assess. 25 (3), 218–235. https://doi.org/10.1080/10627197.2020.1804352.
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. Front. Psychol. 10, 1–32. https://doi.org/10.3389/fpsyg.2019.00646.
Liu, Y., Yang, B., Wu, L., Li, B., Yu, S., 2020. Middle-school students' behavior pattern and strategy selection in problem solving: a study based on data from PISA 2012. In: So, H.J., et al. (Eds.), Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education.
Loh, C.S., 2012. Information trails: in-process assessment of game-based learning. In: Ifenthaler, D., Eseryel, D., Ge, X. (Eds.), Assessment in Game-Based Learning. Springer, Cham, pp. 123–144. https://doi.org/10.1007/978-1-4614-3546-4_8.
Luan, J., 2002. Data mining, knowledge management in higher education, potential applications. In: Workshop Associate of Institutional Research International Conference, Toronto, pp. 1–18.
Mainert, J., Kretzschmar, A., Neubert, J.C., Greiff, S., 2015. Linking complex problem solving and general mental ability to career advancement: does a transversal skill reveal incremental predictive validity? Int. J. Lifelong Educ. 34 (4), 393–411. https://doi.org/10.1080/02601370.2015.1060024.
Markowetz, A., Błaszkiewicz, K., Montag, C., Switala, C., Schlaepfer, T.E., 2014. Psycho informatics: big data shaping modern psychometrics. Med. Hypotheses 82 (4), 405–411. https://doi.org/10.1016/j.mehy.2013.11.030.
Mazza, R., Milani, C., 2005. Exploring usage analysis in learning systems: gaining insights from visualisations. In: Workshop on Usage Analysis in Learning Systems at the 12th International Conference on Artificial Intelligence in Education, pp. 65–72.
McGarr, O., Clifford, A.M., 2013. "Just enough to make you take it seriously": exploring students' attitudes towards peer assessment. High Educ. 65 (6), 677–693. https://doi.org/10.1007/s10734-012-9570-z.
Miller, L.D., Soh, L.K., Samal, A., Kupzyk, K., Nugent, G., 2015. A comparison of educational statistics and data mining approaches to identify characteristics that impact online learning. J. Educ. Data Min. 7 (3), 117–150. https://doi.org/10.5281/zenodo.3554731.
Mislevy, R.J., Steinberg, L.S., Almond, R.G., 2003. On the structure of educational assessments. Measurement 1 (1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02.
Mislevy, R.J., Behrens, J.T., Dicerbo, K.E., Levy, R., 2012. Design and discovery in educational assessment: evidence-centered design, psychometrics, and educational data mining. J. Educ. Data Min. 4 (1), 11–48. https://doi.org/10.5281/zenodo.3554641.
Molnár, G., Csapó, B., 2018. The efficacy and development of students' problem-solving strategies during compulsory schooling: logfile analyses. Front. Psychol. 9, 302. https://doi.org/10.3389/fpsyg.2018.00302.
Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., Heiner, C., 2005. An educational data mining tool to browse tutor-student interactions: time will tell. In: Proceedings of the Workshop on Educational Data Mining, National Conference on Artificial Intelligence. AAAI Press, pp. 15–22.
Mostow, J., 2004. Some useful design tactics for mining its data. In: Proceedings of the ITS2004 Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes, pp. 20–28.
Nicolay, B., Krieger, F., Stadler, M., Gobert, J., Greiff, S., 2021. Lost in transition–Learning analytics on the transfer from knowledge acquisition to knowledge application in complex problem solving. Comput. Hum. Behav. 115, 106594. https://doi.org/10.1016/j.chb.2020.106594.
OECD, 2013. PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy. OECD Publishing, Paris, France. Retrieved from. http://www.oecd-ilibrary.org/education/pisa-2012-assessment-and-analyticalframework_9789264190511-en.
OECD, 2014. PISA 2012 Results: Creative Problem Solving: Students' Skills in Tackling Real-Life Problems (Volume V). OECD Publishing.
OECD, 2017. PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science. OECD Publishing.

Owen, V.E., Ramirez, D., Salmon, A., Halverson, R., 2014. Capturing learner trajectories in educational games through ADAGE (assessment data aggregator for game environments): a click-stream data framework for assessment of learning in play. In: American Educational Research Association Annual Meeting, pp. 1–7.
Paulsen, M.F., 2003. Online Education: Learning Management Systems: Global E-Learning in a Scandinavian Perspective. NKI Forlaget.
Qian, M., Clark, K.R., 2016. Game-based learning and 21st century skills: a review of recent research. Comput. Hum. Behav. 63, 50–58. https://doi.org/10.1016/j.chb.2016.05.023.
Ren, Y., Luo, F., Ren, P., Bai, D., Li, X., Liu, H., 2019. Exploring multiple goals balancing in complex problem solving based on log data. Front. Psychol. 10, 1975. https://doi.org/10.3389/fpsyg.2019.01975.
Romero, C., Ventura, S., 2007. Educational data mining: a survey from 1995 to 2005. Expert Syst. Appl. 33 (1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005.
Romero, C., Ventura, S., 2020. Educational data mining and learning analytics: an updated survey. Wiley Interdiscip. Rev. 10 (3), e1355. https://doi.org/10.1002/widm.1355.
Rupp, A.A., Levy, R., Dicerbo, K.E., Sweet, S.J., Crawford, A.V., Caliço, T., Benson, M., Fay, D., Kunze, K.L., Mislevy, R.J., Behrens, J.T., 2012. Putting ECD into practice: the interplay of theory and data in evidence models within a digital learning environment. J. Educ. Data Min. 4 (1), 49–110. https://doi.org/10.5281/zenodo.3554643.
Rust, J., Golombok, S., 2014. Modern Psychometrics: The Science of Psychological Assessment. Routledge.
Saeed, M., Tahir, H., Latif, I., 2018. Teachers' perceptions about the use of classroom assessment techniques in elementary and secondary schools. Bull. Educ. Res. 40 (1), 115–130.
Salles, F., Dos Santos, R., Keskpaik, S., 2020. When didactics meet data science: process data analysis in large-scale mathematics assessment in France. Large Scale Assess. Educ. 8, 1–20. https://doi.org/10.1186/s40536-020-00085-y.
Sawyer, R., Rowe, J., Azevedo, R., Lester, J., 2018. Filtered Time Series Analyses of Student Problem-Solving Behaviors in Game-Based Learning. Paper Presented at the 11th International Conference on Educational Data Mining (Raleigh, NC, July 16–20).
Scherer, R., Greiff, S., Hautamäki, J., 2015. Exploring the relation between time on task and ability in complex problem solving. Intelligence 48, 37–50. https://doi.org/10.1016/j.intell.2014.10.003.
Schweizer, F., Wüstenberg, S., Greiff, S., 2013. Validity of the MicroDYN approach: complex problem solving predicts school grades beyond working memory capacity. Learn. Indiv Differ 24, 42–52. https://doi.org/10.1016/j.lindif.2012.12.011.
Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., Härtig, H., 2016. Teaching the control-of-variables strategy: a meta-analysis. Dev. Rev. 39, 37–63. https://doi.org/10.1016/j.dr.2015.12.001.
Sharma, K., Papamitsiou, Z., Olsen, J.K., Giannakos, M., 2020. Predicting learners' effortful behaviour in adaptive assessment using multimodal data. In: Proceedings of the 10th International Conference on Learning Analytics & Knowledge, pp. 480–489.
Shute, V.J., Leighton, J.P., Jang, E.E., Chu, M.W., 2016a. Advances in the science of assessment. Educ. Assess. 21 (1), 34–59. https://doi.org/10.1080/10627197.2015.1127752.
Shute, V.J., Wang, L., Greiff, S., Zhao, W., Moore, G., 2016b. Measuring problem solving skills via stealth assessment in an engaging video game. Comput. Hum. Behav. 63, 106–117. https://doi.org/10.1016/j.chb.2016.05.047.
Shute, V.J., 2011. Stealth assessment in computer-based games to support learning. Comput. Games Instr. 55 (2), 503–524.
Sin, K., Muthu, L., 2015. Application of big data in educational data mining and learning analytics—a literature review. ICTAC J. Soft Comput. 5 (4), 1035–1049. https://doi.org/10.21917/ijsc.2015.0145.
Slater, S., Joksimović, S., Kovanovic, V., Baker, R.S., Gasevic, D., 2017. Tools for educational data mining: a review. J. Educ. Behav. Stat. 42 (1), 85–106. https://doi.org/10.3102/1076998616666808.
Srivastava, J., Cooley, R., Deshpande, M., Tan, P.N., 2000. Web usage mining: discovery and applications of usage patterns from web data. ACM SIGKDD Explor. Newsl. 1 (2), 12–23.
Stadler, M., Becker, N., Gödker, M., Leutner, D., Greiff, S., 2015. Complex problem solving and intelligence: a meta-analysis. Intelligence 53, 92–101.
Stadler, M., Hofer, S., Greiff, S., 2020. First among equals: log data indicates ability differences despite equal scores. Comput. Hum. Behav. 106442. https://doi.org/10.1016/j.chb.2020.106442.
Steen-Utheim, A., Hopfenbeck, T.N., 2019. To do or not to do with feedback. A study of undergraduate students' engagement and use of feedback within a portfolio assessment design. Assess Eval. High Educ. 44 (1), 80–96. https://doi.org/10.1080/02602938.2018.1476669.
Stone, E., Wylie, E.C., 2019. Building a validity argument while developing and using an assessment: a concurrent approach for the Winsight® summative assessment. ETS Res. Rep. Ser. 2019 (1), 1–20. https://doi.org/10.1002/ets2.12261.
Teig, N., Scherer, R., Kjærnsli, M., 2020. Identifying patterns of students' performance on simulated inquiry tasks using PISA 2015 log-file data. J. Res. Sci. Teach. 57 (9), 1400–1429. https://doi.org/10.1002/tea.21657.
Trilling, B., Fadel, C., 2009. 21st Century Skills, Learning for Life in Our Times. John Wiley & Sons.
Vainikainen, M.P., 2014. Finnish Primary School Pupils' Performance in Learning to Learn Assessments: A Longitudinal Perspective on Educational Equity. Report 360. Unigrafia, Helsinki, Finland.
Van der Aalst, W.M.P., Pesic, M., Song, M., 2010. Beyond process mining: from the past to present and future. In: International Conference on Advanced Information Systems Engineering. Springer, Cham, pp. 38–52.
Van der Aalst, W.M.P., Guo, S., Gorissen, P., 2015. Comparative process mining in education: an approach based on process cubes. In: Ceravolo, P., Accorsi, R., Cudre-Mauroux, P. (Eds.), Data-Driven Process Discovery and Analysis. SIMPDA 2013, Lecture Notes in Business Information Processing, vol. 203. Springer. https://doi.org/10.1007/978-3-662-46436-6_6.
Van der Aalst, W.M.P., 2012. Process mining: overview and opportunities. ACM Trans. Manag. Inf. Syst. 3 (2), 1–17. https://doi.acm.org/10.1145/0000000.0000000.
Van der Kleij, F.M., Vermeulen, J.A., Schildkamp, K., Eggen, T.J.H.M., 2015. Integrating data-based decision making, assessment for learning and diagnostic testing in formative assessment. Assess Educ. Princ. Pol. Pract. 22 (3), 324–343. https://doi.org/10.1080/0969594X.2014.999024.
Van Groen, M.M., Eggen, T.J., 2020. Educational test approaches: the suitability of computer-based test types for assessment and evaluation in formative and summative contexts. J. Appl. Test. Technol. 21 (1), 12–24.
Von Davier, A.A., Deonovic, B.E., Yudelson, M., Polyak, S., Woo, A., 2019. Computational psychometrics approach to holistic learning and assessment systems. Front. Educ. (4), 69. https://doi.org/10.3389/feduc.2019.00069.
Vrzakova, H., Amon, M.J., Stewart, A., Duran, N.D., D'Mello, S.K., 2020. Focused or stuck together: multimodal patterns reveal triads' performance in collaborative problem solving. In: Proceedings of the 10th International Conference on Learning Analytics & Knowledge, pp. 295–304.
Weeks, J.P., von Davier, M., Yamamoto, K., 2016. Using response time data to inform the coding of omitted responses. Psychol Test Assess. Model. 58, 671–701.
Wijesooriya, C., Heales, J., Clutterbuck, P., 2015. Forms of formative assessment in virtual learning environments. In: 21st Americas Conference on Information Systems, pp. 1–16.
Wüstenberg, S., Greiff, S., Funke, J., 2012. Complex problem solving. More than reasoning? Intelligence 40, 1–14. https://doi.org/10.1016/j.intell.2011.11.003.
Xu, H., Fang, G., Chen, Y., Liu, J., Ying, Z., 2018. Latent class analysis of recurrent events in problem-solving items. Appl. Psychol. Meas. 42 (6), 478–498. https://doi.org/10.1177/0146621617748325.
Yamamoto, K., Lennon, M.L., 2018. Understanding and detecting data fabrication in large-scale assessments. Qual. Assur. Educ. 26, 196–212. https://doi.org/10.1108/QAE-07-2017-0038.

Yang, F., Li, F.W., 2018. Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. Comput. Educ. 123, 97–108. https://doi.org/10.1016/j.compedu.2018.04.006.

Zehner, F., Harrison, S., Eichmann, B., Deribo, T., Bengs, D., Anderson, N., Hahnel, C., 2020. The NAEP EDM competition: on the value of theory-driven psychometrics and machine learning for predictions based on log data. In: Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020), pp. 302–312.

Zheng, G., Fancsali, S.E., Ritter, S., Berman, S., 2019. Using instruction-embedded formative assessment to predict state summative test scores and achievement levels in mathematics. J. Learn. Anal. 6 (2), 153–174. https://doi.org/10.18608/jla.2019.62.11.

Zhu, M., Shu, Z., von Davier, A.A., 2016. Using networks to visualize and analyze process data for educational assessment. J. Educ. Meas. 53 (2), 190–211. https://doi.org/10.1111/jedm.12107.

Zoanetti, N., Griffin, P., 2017. Log-file data as indicators for problem-solving processes. In: Csapó, B., Funke, J. (Eds.), The Nature of Problem Solving: Using Research to Inspire 21st Century Learning. OECD Publishing. https://doi.org/10.1787/9789264273955-13-en.