The Invisible Breadcrumbs of Digital Learning:
How Learners Actions Inform Us of Their Experience

Introduction

One of the main opportunities provided by digital learning is the capability not only to examine the final products of learning activities (e.g., essays, final answers to problems and test scores), but also to collect detailed logs of how learners participate in learning activities. These logs can be especially valuable when digital learning takes place outside of a traditional classroom, where it would be difficult or impossible for an instructor to monitor learners as they engage with their learning activities—for example, when completing after class homework or when participating in asynchronous online courses.

Action logs consist of records of all the actions that the learners execute within a learning environment. These action logs are comparable to invisible breadcrumbs left behind by learners, marking the path they took as they engaged with the environment and providing fine-grained information about when and how they interacted with specific components of the environments' user interface. Studying and examining these breadcrumbs can be a valuable source of information, allowing us to follow the learners through their learning experiences in order to better understand and support them.

This chapter discusses how logs of learners' actions within digital learning environments can provide us with information about their learning experiences. It starts by describing in more detail what action logs are and what kind of information they contain. Then, it discusses two research communities, Educational Data Mining (EDM) and Learning Analytics (LA), who are particularly interested in the usage of action logs to study learning and digital learning environments. This is followed by selected examples of how action logs can be analyzed and how the result of those analyses can be used to support learning.

Action logs

A unique aspect of digital learning environments is their capability to collect detailed logs of the learners' actions. Generally, such logs will contain a list of the meaningful actions that each learner executed within the digital learning environments, accompanied by the exact times at which each action occurred and any other important variables that are needed to describe the action that occurred. As each digital environment allows for different sets of meaningful action, the format and content of action logs can greatly vary between the different environments. To better illustrate what type of information interaction logs can contain, three examples, from three different types of digital learning environments, are presented and an explanation of what their action logs might contain is provided: Cognitive Tutor Algebra (Koedinger & Corbett, 2006), an intelligent tutoring system in which learners solve mathematical problems; Physics Playground (Shute et al. 2013), an educational game in which learners construct simple machines and use principles of Newtonian physics in order to guide a ball to a balloon; and massive open online course (MOOC) platforms such as Coursera[1] and EdX[2].

Example #1: Cognitive Tutor Algebra

Cognitive Tutor Algebra (Koedinger & Corbette, 2006) is an intelligent tutoring system, in which learners practice solving algebra problem. Solving a problem within Cognitive Tutor Algebra requires learners to complete multiple steps towards computing the final answer. The tutor evaluates the correctness of each step as soon as the learner completes them, identifying whether a step is correct, contains a known error (called a bug), or is otherwise incorrect, and provides

[1] https://www.coursera.org/
[2] https://www.edx.org/

immediate feedback to the learner. In addition, at any moment learners have the option of requesting hints about how to execute the next step of the problem-solving process.

As the number of types of actions that can be executed in Cognitive Tutor Algebra is very limited—an action can either be a step or a hint request—the associated action logs can be kept simple. Each entry in the log will provide information about one action, its type (step attempt or hint request), the exact time at which the action happened and, if the action is a step attempt, the name of the step that was attempted and how the tutor evaluated it (correct, bug, or incorrect). This simple format allows for the reconstruction of a learner's attempt at solving a problem within Cognitive Tutor Algebra.

Example #2: Physics Playground

Learners mainly interact with Physics Playground (Shute et al. 2013) by drawing simple physical machines (e.g., ramps, pendulums, levers and springboards) to solve physics-based puzzles in which they need to guide a ball towards a balloon. While they play the game, the motion of any every object within the puzzle (including the ball) is simulated using the basic laws of physics. In addition, learners can pause and restart the simulation when desired.

Since Physics Playground provide a larger variety of actions and a more dynamic environment than intelligent tutoring systems such as Cognitive Tutor Algebra, the action logs it produces are also more complex. Simple actions, such as pauses and restarts of each puzzle, can be described by providing the time at which the action occurred. However, other actions—for example, a click on a particular point on the screen—might require additional information, such as the 2D coordinate at which the click occurred. Additionally, complex actions such as drawing a machine will include logs of which type of machine was drawn and provide a detailed list of coordinates that were used when drawing it. In addition to logging the actions executed by the learners, educational games such as Physics Playground can also log meaningful events that are useful to the interpretation of the learners' actions. For example, an event can be logged when two objects collide with each other—providing information about when the collision occurred and which objects where involved—or when an object, such as one of the drawn machines or the ball, leaves the play area.

Example #3: MOOCs

Although the content and the structure of each MOOC is unique, MOOC platforms usually provide learners with a similar set of tools. A MOOC will be composed of a set of webpages organizing pedagogical content for the course. Generally, instruction is delivered through a set of videos. Learners can interact with each other in discussion forums, and learning is assessed through quizzes and assignments.

In this context, action logs will mostly consist of information about when learners access pedagogical content and how they interact with this content. This includes information about which webpage was accessed and when. Video watching actions can also be included in the logs, providing information not only about which video was played, but also how the learner interacted with the video by changing playback speed, pausing/unpausing, or seeking through the video. Similarly, information can be collected about which discussion forums learners accessed and what text they post in the forums. Logs related to quizzes and assignments can provide information about which assessment was attempted, how often, what answers were provided, and what the learners' scores were.

Why are action logs informative?

Actions logs can be a rich source of information to better understand the thought processes of learners as they engage with digital learning environments. Each action produced within the learning environment can be recorded and interpreted, similar to a teacher or a tutor observing a learner as they engage with more traditional learning tasks. Analyzing these actions thus affords a better understanding of the learners' knowledge and their strategies when approaching the task, in order to better support their learning experience. The more detailed the action logs and the more traces left in the learning environment by the learner, the more informative action logs can be.

For example, within intelligent tutors such as Cognitive Tutor Algebra, pauses between actions can be interpreted in conjunction with those actions to attempt to infer the learner's problem-solving approach. A short pause before attempting to complete a step of a problem could be an indicator of mastery of the material (if the attempt is successful), or of a lack of forethought if the attempt is unsuccessful. Similarly, a long pause can be an indicator of a genuine attempt at understanding how to complete the step. When requesting hints, the length of a pause can be used to infer whether the learner took the time to read and understand the instruction that were provided to them. Looking at sequences of actions through such a lens can provide insights about the thought process and problem-solving strategies that were used by learners when solving problems.

Although they can be very informative, it is important to consider that action logs are an incomplete source of information about the learners' thought processes. Logs only include actions that have been executed within the digital learning environment and fail to capture any activity outside of it (e.g., writing on a sheet of paper or discussing the tasks with other learners). Additionally, it is important to consider that multiple interpretations of the learners' thought process could lead to the same action within the learning environment. As such, interpreting action logs requires careful consideration of those multiple possible interpretations, and the usage of methodologies designed to accumulate evidence to identify the most probable interpretations.

Educational Data Mining and Learning Analytics

Multiple research communities have taken a recent interest in studying how we can make sense of action logs to better inform us of learners' experiences and how we can provide new opportunities to better support them as they engage with digital learning environments. In particular, both the Educational Data Mining (EDM) and the Learning Analytics (LA) research communities study students' action logs to improve student learning and to discover more about the learning process itself. Broadly speaking, the EDM community (Baker & Yacef, 2009) focuses on theory and applications of data mining approaches that are suited to the kinds of data that commonly result from recorded learning experiences. For example, these data could be videos of classrooms (Raca et al. 2015), audio recordings of teachers speaking (Blanchard et al. 2015), or logs of students' actions in digital learning environments—each of which presents challenges that are often unique to education. The LA community (Siemens & Long, 2011) typically follows a similar approach to research and the types of data that are analyzed but places an emphasis on research covering so-called "meta-issues" of data-driven analyses in education. These issues include, for example, ethical and privacy implications of data mining, student perspectives about these analytics, and strategies for encouraging wider adoption of LA methods in education.

EDM and LA communities approach the problem of modeling learning experiences from complementary angles of educational theory, human psychology, and computer science (Siemens & Baker, 2012). Theory informs the design of analyses and digital learning environments (e.g., suggesting what types of interventions are likely to be successful in adaptive learning software), while data mining helps to answer theoretical questions (e.g., what types of strategies do students take when approaching new learning materials).

Action log analysis has evolved to become a common component of both theoretical and applied studies, particularly with the emergence of MOOCs and other web-based learning environments. In these environments there are few opportunities for studying the learning process, given that students are typically not physically near teachers or researchers who can observe them. However, action logs provide a window through which students can still be observed and studied. The EDM and LA research communities have evolved along with these changes in how students learn. For example, in the first edition of the Learning Analytics & Knowledge conference (2011) there were no publications that mentioned MOOCs in the title, while in eighth year (2018) there were two entire presentation sessions devoted to aspects of MOOC analysis. A similar trend holds true for the Educational Data Mining conference.

Clearly, a variety of different data types (not just action logs) can provide valuable insights into learning, and in turn improve students' learning experiences. Students' facial expressions can convey their engagement with learning material, as can their gaze direction. Researchers in the EDM and LA communities often consider these types of data as well, but action logs are more commonly available since they do not require any specific equipment (e.g., webcam, eye-tracker). Thus, their analysis is becoming increasingly important as digital and web-based learning gains popularity.

Analyzing action logs

Once action logs have been collected for a digital learning environment, approaches used and developed by the EDM and LA communities can be used to make sense of learners' actions and better understand their thought processes. This section briefly discusses three broad categories of analyses that can be conducted using action logs to better understand learners' experiences. It is important to note that these three categories do not constitute a complete list of all the analyses that action logs allow, but rather are selected as a sample of typical approaches used to take advantage of action log data.

Descriptive and exploratory analytics

Action logs can be used to provide descriptive and exploratory analytics, in which logs are aggregated across multiple learners or contexts to produce a summary of how learners interact with the digital learning environment. This type of analysis does not attempt to infer the thought process behind the learners' actions, but rather has the goal of generally describing how learners use the environment. Such analytics can provide a better understanding of the learners' actions or lead to important questions about why learners interact with the environment in the observed way.

In the context of MOOCs, descriptive analytics can illustrate how learners interact with instructional material such as videos. For example, some researchers have used descriptive analytics to study how learners engage with instructional video and what can be learned from those engagement patterns (Giannakos et al. 2015; Guo et al. 2014; Kim et al. 2014). While watching a video is a mostly passive activity, learners can interact with videos by executing actions such as seeking forward or backward, pausing the video or resuming it. Kim et al. (2014) leveraged descriptive analytics to study the time at which learners perform those actions during video playback. They discovered the presence of moments of increased activity, or peaks in video interaction, suggesting points of interest in the videos and raising questions about the thought processes that might lead to increased interaction with those points of interest. They identified five types of activities related to those peaks: "starting from the beginning of a new material, returning to missed content, following a tutorial step, replaying a brief segment, and repeating a non-visual explanation" (Kim et al. 2014).

Another example of the usage of descriptive analytics and how such analytics can lead to meaningful observations related to how students learn from digital environments can be observed in the work by Bosch & D'Mello (2017). In their study, they examined sequences of actions taken by students learning computer programming for the first time in a digital learning environment. Students in this environment were able to write code, test their code, and then submit it to have it automatically graded. The researchers found that when students spent some time coding, then tested their code, and encountered an error in the test, they would return to revising their code. However, if students submitted their code to be graded and encountered an error, they tended to request a hint—suggesting that this type of error indicates a student who is at an impasse and needs guidance rather than simply more time revising their code. These analyses were purely descriptive, but lead naturally to the next type of analysis discussed in this chapter, in which researchers attempt to explicitly connect learners' actions to specific behaviors they are interested in.

Behavior modeling

Unlike descriptive analytics, behavior modeling attempts to make inference about the learners' behaviors based on the actions they performed. Common uses of behavior modeling include associating actions to specific known behaviors (often using supervised machine learning approaches, such as classification algorithms, to build models able to detect the behavior), and discovering previously unknown or unanticipated behavior patterns in the learners' actions (often using unsupervised machine learning approaches, such as clustering).

In the context of intelligent tutoring systems, for example, behavior modeling has been used to identify learners who misuse the learning environment by systematically attempting to guess answers (Baker et al. 2008; Paquette et al. 2014) or by abusing hint request functionalities (Aleven et al. 2006). Requesting hints is an integral part of how students learn within intelligent tutoring systems. When used appropriately, hints provide instructions upon which learners can reflect to acquire new knowledge about how to solve problems. However, hints can also be abused by requesting too many hints with the goal of having the tutor provide the answer. Action logs can help distinguish between productive and abusive uses of hint requests. For example, pauses between hint requests can be an especially good indicator of how the learners use hints. Aleven et al. (2006) have identified a list of inappropriate hint usages that can be detected using action logs, including quickly clicking through hints without taking time to read them. Action logs can also be useful to identify appropriate behavior resembling inappropriate ones. For example, Shih et al. (2011), have observed that, although quickly clicking through hints is mostly inappropriate (identified as “quickly clicking through hints by Aleven), such behavior can be useful for learners who use those hints as worked examples—taking long pauses after the hint requests to understand how to reproduce the answer given by the tutor. Paquette et al. (2014) also studied how sequences of actions and pauses within intelligent tutoring systems can be leveraged to identify learners who misuse intelligent tutors, for example, by attempting to systematically guess answers.

In contrast to using action logs to model specific known behaviors, unsupervised machine approaches (e.g., clustering algorithms) can be used to search for patterns that naturally emerge from the logs themselves. Such an exploratory analysis can be used to identify common strategies that learners adopt when using the digital learning environment. For example, Li et al. (2015) used clustering to find common behaviors of learners when watching instructional video. In their study, they calculated eight indicators of how learners interact with videos during playback (number of pauses, median pause duration, number of forward seeks, proportion of skipped content, number of backward seeks, length of replayed video, average video speed and effective video speed change). Clustering analysis allowed them to group video watching behaviors together based on how similar they were to each other (when comparing the values of those eight indicators). Using this method allowed them to identify nine distinct behaviors that can be used to better understand how learners approach video watching tasks and to study the quality of those videos: long replays, high-speed playback, speeding up the video, skimming and skipping, inactive, frequent pauses, jump skip, long pauses and slowing down the video.

Knowledge estimation

Another use for action logs is to infer what a learner knows and does not know. This can be achieved by observing successful and unsuccessful attempts at using specific skills and concepts related to a topic they are learning. For example, to infer whether a learner knows how to add two single-digit numbers together, we can wait for the learner to attempt such an addition and observe its outcome. A successful addition would suggest that the learner knows how to add two numbers and an unsuccessful attempt would suggest that the learner is lacking this knowledge. However, observing a single successful (or unsuccessful) attempt is rarely sufficient to infer knowledge (or lack of knowledge). Instead, actions logs can be used to accumulate observations over time and look at patterns of successful and unsuccessful attempts, with each observation of success (or failure) providing additional evidence of knowledge (or lack thereof).

A common approach for the estimation of knowledge in intelligent tutoring systems is knowledge tracing, and most commonly Bayesian Knowledge Tracing (BKT; Corbett & Anderson, 1995). In BKT every action that is associated with a specific skill is considered as an opportunity to acquire this skill, and every past success and failure to correctly apply this skill is considered as evidence to estimate the learner's knowledge. Each observed success increases the likelihood that the learner mastered the skill in question, whereas each failure decreases it. In addition, knowledge tracing algorithms often account for the possibility that a learner will fail to successfully apply a skill that they have acquired—traditionally called a "slip"—or that they will successfully answer a question that requires a skill that they have not yet mastered—traditionally called a "guess".

Although BKT is the most popular knowledge estimation approach for intelligent tutors, it is not the only one. Multiple approaches have been developed using different methods such as Performance Factor Analysis (Pavlik et al. 2009), using a logistic regression modeling approach, and Deep Knowledge Tracing (Piech et al. 2015), using a deep neural network modeling approach. Additionally, applications of knowledge estimation are not limited to intelligent tutors, as knowledge tracing algorithms have been applied to multiple types of digital learning environments including educational games (Gweon et al. 2015), MOOCs (Pardos et al. 2013), and online resources such as Khan Academy (Piech et al. 2015).

Using action logs to support learning

The previous section provided examples of how action logs can be analyzed to provide insights about learners' thought processes and about their learning experience. Those examples are not meant to be an exhaustive list of ways in which analyses of action logs can be used to support learning, but rather a set of important themes. Furthermore, although those analyses can be useful by themselves when studying the process of learning, it is also important to consider how the results of those analyses can be leveraged to create learning environments that can better support the learning process. This section provides examples of three ways in which the results of action log analyses can be used to support learning: data-driven design, informing instructors, and automatically adapting the learning environment.

Data-driven design

Analyses of the action logs can provide useful information about how the learners use the digital learning environment, providing insights about what seems to be working well, what they are struggling with, and if they behave in unexpected ways. Those insights can then be used to rethink the design of the learning environment or to identify best practices for future designs.

In the context of online courses, data-driven design has proven to be a valuable way to benefit from action log analyses (Guo et al. 2014). Since online courses, especially MOOCs, are often provided in an asynchronous format, it can be difficult to obtain direct information about how the learners' experience the courses' instructional material. Surveys can be used to directly ask learners to provide their opinion, however, they usually have very low response rates (e.g., 11% response rate in Crues et al., 2018) and are susceptible to self-selection bias regarding who completes them. Alternatively, action log analyses—like those described in the previous section—can be used as an indirect measure of the learners' experiences with the courses material. For example, Guo et al. (2014) compared how engaging different formats of instructional videos were by looking at how long learner spent watching each video and whether they answered post-video assessment. By comparing those two measures across a large sample of learners and videos, the researchers were able to identify trends regarding what types of video were most engaging, and they subsequently developed a set of design recommendations. They recommend, among other things, short videos (no more than 6 minutes), filming in informal contexts, and editing videos to display the instructors head at opportune times—rather than designing videos that only convey information through slideshows.

Action log analyses can also inform the design of digital learning environments throughout the different phases of their development. Owen (2015) proposed a data-driven design approach for educational game development which integrates analytics in the early core design phase, the early development phase, and the final stages of design. Owen explains how, during the core design phase, carefully thinking about the data framework that will be used to log learners' actions can contribute to the design of the game itself. Considering what data can and should be collected allows the designers to think about how the actions within the game are linked to the learning experience and how actions can serve as evidence of learning. During the early development, action logs, combined with descriptive analytics and visualizations of those analytics, can be particularly useful. For example, descriptive analytics regarding how often learners access different game levels might reveal issues in the game's interface where some levels are difficult to discover (Beall et al. 2013). Similarly, using heatmaps to visualize how learners navigate the game's environment allows for the discovery of frequently visited areas and can inform the placement of critical in-game assets. Finally, during the final stages of design, more comprehensive analytics can be used to model and predict the learners' actions and performance within the game, such as classification modeling, association rule mining, or sequential pattern analysis. By identifying actions and events that are most indicative of learning, designers can develop in-game scaffolding to support learners at key events in the game.

Informing instructors

Action logs can also be used to empower instructors by informing them of the learners' activities within digital learning environments. This type of support is especially useful when learners engage in activities outside of the classroom or when it is difficult for the instructor to pay close attention to how the learners interact with the learning environment. In such situations, action logs can be leveraged to produce real-time or retrospective reports summarizing information about the learners' activities, to inform instructors in a format that is easily interpretable and useful to them.

An example of how action logs can be used to inform instructors can be seen in the work by Kay and colleagues on the visualization of collaborative processes in digital learning. With her team, Kay has developed multiple ways to visualize learner behavior as they interact with multiple collaborative tasks over the short and long term, including semester-long software development projects (Kay et al. 2006) and shorter activities where learners engage in the creation of a concept map using interactive tabletop computers (Martínez et al. 2011; Maldonado et al. 2012). They integrated their visualization of tabletop usage in an interactive dashboard tool designed to provided instructors with a way to monitor the learners' progress towards building their concept map and their collaboration with each other. Using this tool, instructors engaged in a classroom orchestration loop (Dillenbourg et al. 2011) in which they used the visualizations to identify groups with unexpected collaboration patterns, acquire information about ways in which their collaboration could be supported, and provide feedback to the group. Beyond the work by Kay and colleagues, there has been a growing interest in the design of dashboards that take advantage of action log analytics to support instructors and students (Verbert et al. 2013) across diverse types of digital learning environments, including e-learning systems (France et al. 2006; Podgorelec & Kuhar, 2011) and intelligent tutoring systems (Holstein et al. 2017).

Automatic adaptation

The two previously-described categories of approaches to using action logs to support learners have focused on giving back the information collected from action logs to relevant stakeholders (e.g., designers or instructors) to support them. Another way to take advantage of action logs is to develop digital learning environments that make use of this information by dynamically adapting content or instructional methods based on the learners' actions. The goal of this approach is to provide the learners with an experience that is personalized to their own needs.

One way for intelligent tutoring systems to personalize their instruction is to automatically select what problem the learner should solve next. The goal is to select a problem that will be of appropriate difficulty for the learner, avoiding problems that are too easy and would not be a beneficial use of the learner's time, as well as problems that would be too difficult for the learner to successfully complete. Probably the most famous of automatic example of automated problem selection in an intelligent tutor comes from the Cognitive Tutors (Anderson et al. 1995), where they took advantage of knowledge tracing (described in the previous section) to implement a mastery learning approach (Bloom 1984) within their tutoring system. In mastery learning, the learner repeatedly attempts to solve similar problems until they have reached a satisfactory level of mastery before moving on to new and more difficult problems. Knowledge tracing allows the tutor to track what skills the learner masters and which ones still need to be learned. To choose an appropriate problem, the tutor can compare the list of skills required to solve a specific problem to the list of skills for which the learner has achieved mastery. A problem containing only mastered skills would be classified as being too easy, whereas a problem containing mostly skills that have yet to be mastered would be too difficult. Instead, the tutor will attempt to find a problem that provides a good balance between skills that have already been mastered and skills that have not yet been mastered but that the learner should be ready to acquire.

The learning environment can also automatically adapt to specific behaviors that the learners engage in. For example, Baker and colleagues (2006) designed interventions to mitigate the issues that occur when students misuse a Cognitive Tutor. As described above, action logs can be used to build models that detect when students engage in such behavior. This allows the learning environment to automatically provide remedial support when a learner is detected as misusing the environment. Since the learner is avoiding engaging in learning opportunities by misusing the environment, Baker and colleagues (2006) designed a tutor that would automatically provide additional conceptual questions (in contrast to the usual problem-solving problems provided by the tutor) specifically targeting the concepts that were missed due to misuse. Those questions provided the learners with additional opportunities to learn the concepts that might have been more difficult for them and might have led to more misuse.

Concluding remarks

The aim of this chapter was to bring attention to action logs and the key role they can play in digital learning. Unlike the more explicit outcomes of learning activities—e.g., essays, final answers to problems and test scores—action logs are less visible and, as such, are more likely to be ignored. However, they provide rich information as they not only inform us about the end results of the learning process, but also about the learners' thought processes while they were actively engaged in the learning activities. In this regard, action logs can be compared to a trail breadcrumbs left behind to mark the path taken by learners and allowing us to retrace this path to better support them.

Diverse approaches can be used to analyze action logs to getter a better understanding of the learners' thought process and their experiences with digital learning environments. Each approach brings a distinct perspective, ranging from more general observations that can be made available using exploratory and descriptive analytics to the identification of more specific learning behaviors and strategies or the estimation of the learners' knowledge. The result of those analyses can in turn be used to better support the learners through their learning process—for example, by using data-driven design to improve the learning environment itself, informing instructors about the learners' experience, or automatically adapting the learning environment to the learners' need.

Despite action logs being a great resource to study and support the learning process, it is important to be mindful of some of the issues surrounding action log analysis. In particular, it is crucial that they are used in a responsible and meaningful way, especially in regards with working with uncertainty and ensuring ethical usage of the action logs.

Working with uncertainty

Although learners' actions within a digital learning environment provide us with valuable insights about their thought processes and their learning experience, it is important to remember that action logs are an incomplete source of information. Analyses conducted on action logs and models created from them can reveal common trends but might not be accurate in every possible situation. Analyses allow us to generate hypotheses about the learners' thought processes, accumulate evidence in support of hypotheses, and make inferences based on well-supported hypotheses. However, many factors can influence how learners act when using digital learning environments. As such, it is unrealistic to expect models developed from action logs to be able to perfectly identify the right hypothesis and make the best inference in every situation. This is especially true when conditions in one environment differ notably in some way from the environment in which data were originally collected and analyses were performed (e.g., a classroom versus a laboratory environment).

Such inaccuracies need to be carefully considered when conducting analyses of action logs, interpreting their results and considering how to apply those results to support learning. The result of analyses need to be validated across multiple learners or populations of learners to ensure that they generalize equally well for everyone, they need to be used in a way that is going to be beneficial for everyone, and researchers need to ensure that they will not cause any harm—even in situations where the models make inaccurate inferences.

However, models need not be perfect to be effective tools in digital learning environments. In some cases, a moderately accurate model of student behavior is just as beneficial for aiding learning as a perfectly accurate model. For further reading on this topic, Kitto and colleagues

(2018) provide a detailed discussion on working with and embracing imperfection in LA and EDM research.

Ethical use of action logs

In addition, it is important to consider the ethical issues (Slade & Prinsloo, 2013) of using action logs. As mentioned in the previous section, using models derived from action logs will most likely require working with uncertainty. This implies that, although interpretation of the analysis will hold true on average, it does not automatically mean that the results hold true for every individual learner. As such, it is important to consider the ethical implications of using the results of action log analyses to drive digital learning environment. The resulting design changes (e.g., interventions, automatic adaptations) might be beneficial for most students, but researchers must carefully avoid cases in which changes might be ineffective or, in the worse-case scenario, even harmful. For example, what if the results of the action log analysis are inadvertently biased for or against a specific population of learners? It is our responsibility to ensure that our applications of action log analyses consider the underlying ethical implications and always strive to maximize benefits while avoiding or, at the very least minimize, potential harm.

A second critical issue is that of respecting the learners' privacy (Zeide 2017). Unlike with essays or other tests that learners explicitly turned in as a result of their learning activities, action logs are automatically collected—often without the learners' knowledge. As such, learners often have no direct control over the data being collected. It is our ethical responsibility to ensure that any collection or analysis of action logs respect the learners' privacy. This includes consideration such as whether the digital learning environment should explicitly disclose that it is collecting action log data, where the collected data is stored and who has access to it, and whether the collected data includes any personally identifiable information about the learners.

These ethical implications are not unique to action log analysis; for example, digital learning environments that adapt to student attention via eye-gaze tracking face similar issues of ensuring fairness for all students and preserving privacy. Thus, similar approaches can be adopted for action log analyses—such as obtaining informed consent from learners whose data is being collected and collecting demographic information about learners to quantify differences in analyses between groups of students. With these considerations in mind researchers can leverage action logs to their fullest potential, creating a better learning experience for all students.

References

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education*, *16*(2), 101-128.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*(2), 167-207.

Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R., Evenson, S., Roll, I., Wagner, A. Z., ... & Beck, J. E. (2006). Adapting to when students game an intelligent tutoring system. In *International Conference on Intelligent Tutoring Systems* (pp. 392-401). Springer, Berlin, Heidelberg.

Baker, R. S. J. d., Corbett, A. T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction*, *18*(3), 287-314.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM Journal of Educational Data Mining*, *1*(1), 3-17.

Blanchard, N., D'Mello, S. K., Olney, A. M., & Nystrand, M. (2015). Automatic classification of question & answer discourse segments from teacher's speech in classrooms. In

Proceedings of the 8th International Conference on Educational Data Mining. International Educational Data Mining Society.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4-16.

Bosch, N., & D'Mello, S. (2017). The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education*, *27*(1), 181-206.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-adapted Interaction*, *4*(4), 253-278.

Crues, R. W., Bosch, N., Anderson, C. J., Perry, M., Bhat, S., & Shaik, N. (2018). Who they are and what they want: Understanding the reasons for MOOC enrollment. In *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 177-186). International Educational Data Mining Society.

Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., Bonnard, Q., Cuendet, S., & Kaplan, F. (2011). Classroom orchestration: The third circle of usability. In *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning* (pp. 510-517). International Society of the Learning Sciences.

France, L., Heraud, J. M., Marty, J. C., Carron, T., & Heili, J. (2006, July). Monitoring virtual classroom: Visualization techniques to observe student activities in an e-learning system. In *Proceedings of the Sixth International Conference on Advanced Learning Technologies* (pp. 716-720). IEEE.

Giannakos, M. N., K. Chorianopoulos, & N. Chrisochoides (2015). Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. *The International Review of Research in Open and Distributed Learning*, *16*(1), 260-283.

Guo, P. J., J. Kim, & R. Rubin (2014). How video production affects student engagement: an empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning @ Scale* (pp. 41-50). ACM.

Gweon, G. H., Lee, H. S., Dorsey, C., Tinker, R., Finzer, W., & Damelin, D. (2015). Tracking student progress in a game-like learning environment with a monte carlo bayesian knowledge tracing model. In *Proceedings of the Fifth International Conference on Learning Analytics & Knowledge* (pp. 166-170). ACM.

Holstein, K., McLaren, B. M., & Aleven, V. (2017). Intelligent tutors as teachers' aides: Exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 257-266). ACM.

Kay, J., Maisonneuve, N., Yacef, K., & Reimann, P. (2006). The big five and visualisations of team work activity. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (pp. 197-206). Springer, Berlin, Heidelberg.

Kim, J., S. Li, C. J. Cai, K. Z. Gajos, & R. C. Miller (2014). Leveraging video interaction data and content analysis to improve video learning. In *Proceedings of the CHI 2014 Learning Innovation at Scale workshop*. Toronto, Canada.

Kitto, K., Shum, S. B., & Gibson, A. (2018). Embracing imperfection in learning analytics. In *Proceedings of the 8th Learning Analytics & Knowledge Conference* (pp. 451-460). ACM.

Koedinger, K.R, & Corbett, A.T. (2006). Cognitive Tutors: Technology bringing learning sciences to the classroom. In R.K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 61-77). Cambridge University Press, New York, NY, US.

Li, N., Kidziński, Ł., Jermann, P., & Dillenbourg, P. (2015). MOOC video interaction patterns: What do they tell us?. In *Design for teaching and learning in a networked world* (pp. 197-210). Springer, Cham.

Martínez, R., Collins, A., Kay, J., & Yacef, K. (2011). Who did what? Who said that?: Collaid: an environment for capturing traces of collaborative learning at the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (pp. 172-181). ACM.

Maldonado, R. M., Kay, J., Yacef, K., & Schwendimann, B. (2012). An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (pp. 482-492). Springer, Berlin, Heidelberg.

Owen, V. E. (2015). Using learning analytics to support educational game development: A data-driven design approach. In *American Educational Research Association (AERA) annual meeting*. American Educational Research Association, Chicago, IL.

Paquette, L., de Carvalho, A. M., & Baker, R. S. (2014). Towards understanding expert coding of student disengagement in online learning. In *Proceedings of the 36th Annual Cognitive Science Conference* (pp. 1126-1131). Cognitive Science Society.

Pardos, Z. A., Bergner, Y., Seaton, D. T., & Pritchard, D. E. (2013). Adapting bayesian knowledge tracing to a massive open online course in edX. In *Proceedings of the 6th*

International Conference on Educational Data Mining (pp. 137-144). International Educational Data Mining Society.

Pavlik Jr, P. I., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis--A new alternative to knowledge tracing. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 531-538). Springer, Berlin, Heidelberg.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems* (pp. 505-513). Curran Associates, Inc.

Podgorelec, V., & Kuhar, S. (2011). Taking advantage of education data: Advanced data analysis and reporting in virtual learning environments. *Elektronika ir Elektrotechnika*, *114*(8), 111-116.

Raca, M., Kidzinski, L., & Dillenbourg, P. (2015). Translating head motion into attention - Towards processing of student's body-language. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 320–326). International Educational Data Mining Society.

Shih, B., Koedinger, K. R., & Scheines, R. (2011). A response time model for bottom-out hints as worked examples. In *Handbook of Educational Data Mining* (pp. 201-212). CRC Press.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research*, *106*(6), 423-430.

Siemens, G., & d Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254). ACM.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, *46*(5), 30.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500-1509.

Zeide, E. (2017). Unpacking Student Privacy. In *Handbook of Learning Analytics* (pp. 327-335). Society for Learning Analytics Research.