

A Novel Method for the In-Depth Multimodal Analysis of Student Learning Trajectories in Intelligent Tutoring Systems

Ran Liu¹, John Stamper^{2*}, Jodi Davenport³

Abstract

Temporal analyses are critical to understanding learning processes, yet understudied in education research. Data from different sources are often collected at different grain sizes, which are difficult to integrate. Making sense of data at many levels of analysis, including the most detailed levels, is highly time-consuming. In this paper, we describe a generalizable approach for more efficient yet rich sensemaking of temporal data during student use of intelligent tutoring systems. This multi-step approach involves using coarse-grain temporality — learning trajectories across knowledge components — to identify and further explore “focal” moments worthy of more fine-grain, context-rich analysis. We discuss the application of this approach to data collected from a classroom study in which students engaged in a Chemistry Virtual Lab tutoring system. We show that the application of this multi-step approach efficiently led to interpretable and actionable insights while making use of the richness of the available data. This method is generalizable to many types of datasets and can help handle large volumes of rich data at multiple levels of granularity. We argue that it can be a valuable approach to tackling some of the most prohibitive methodological challenges involved in temporal learning analytics.

Notes for Practice

- Educational software automatically logs student entries, clicks, and other submitted activity. Researchers can use these detailed logs to infer information about student learning processes. However, software cannot log some important events that happen during student learning experiences
- We collected video and audio while students used educational software in the classroom and developed a tool to combine the software data logs with the audio/video data to discover insights about student learning processes.
- The novel STREAMS tool allows researchers to extract audio and video data aligned with features extracted from software logs that reveal unique insights about students’ conceptual struggles and learning trajectories.

Keywords

Learning trajectories, learning curves, intelligent tutoring systems, science learning, multimodal data, mixed methods

Submitted: 02/13/17 — **Accepted:** 11/08/17 — **Published:** 04/09/18

¹Email: ranliu@gmail.com Address: MARI, LLC, 100 N. Pitt St. Suite 240, Alexandria, VA 22314, ORCID ID: 0000-0002-2832-5216

^{*}Corresponding author. ²Email: jstamper@cs.cmu.com Address: Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213

³Email: jdavensp@wested.org Address: WestEd, 730 Harrison St., San Francisco, CA 94107, ORCID ID: 0000-0001-9091-9616

1. Introduction

Temporal analyses are critical to understanding learning processes, yet they are understudied in education research (Barbera, Gros, & Kirschner, 2015; Mercer, 2008). As learning builds on prior knowledge and past experiences, understanding the context and sequence of an intelligent tutoring activity provides important insights into whether and how students learn the content. Students may struggle to learn new content for a variety of reasons that a temporal analysis can help discover. For instance, students who have particular types of erroneous understandings may demonstrate signature patterns of responses or may be more likely to misinterpret particular types of feedback. Further, temporal analyses show how particular types of learning activities can help students better absorb content (e.g., preparation for future learning; Schwartz & Martin, 2004), or

how misunderstandings of initial explanations can create challenges for later learning (e.g., in chemistry, Taber, 2013; physics, Kubricht, Holyoak, & Lu, 2017; and math, Booth, Newton, & Twiss-Garrity, 2014).

With the increasing ubiquity of intelligent tutoring systems (Van Lehn, 2006) in the classroom, large amounts of temporally rich learning process data are readily available through automatic software usage logs. At the same time, advances in educational data mining and learning analytics promise new and highly automated tools for analyzing these data (Koedinger et al., 2015). Ideally, these analyses will create a rich picture of student knowledge and learning processes as they unfold across time (e.g., Graesser, Conley, & Olney, 2012).

Software-based “log data” cannot capture all important temporal learning phenomena, however, and automated data mining and analytic techniques often miss out on insights that contextually rich, multimodal data streams can provide. Even if log data could technically capture all the details of student activity, the large volume of data in non-intuitive formats requires preprocessing and up-front filtering and/or feature selection. Often interesting learning phenomena gets filtered out or goes unnoticed even when it may be technically captured in the log data.

Integrating quantitative and qualitative data provides a much richer understanding of learning trajectories in intelligent tutoring systems. Quantitative log data is limited by the inherent constraints of the system. That is, the design choices limit the range of inputs and determine what data are recorded in the files; the data provides a record of “what” occurred, but not necessarily “why” or “how” the data were generated. Qualitative data, such as video, has far fewer constraints and can provide insights to system designers about unintended student behaviours that result from constraints in a system. For instance, if a field is programmed to limit the number of possible characters, the data captured by log files alone will not provide an accurate representation of a student’s intent in responding.

Context-rich, multimodal data can enhance researchers’ understanding of log data by readily providing concrete examples of student problem-solving that make student struggles and strategies immediately interpretable. In addition, researchers sometimes misunderstand the data they are working with when doing secondary analyses on isolated log data. The integration of multimodal data streams helps to bring the experience of students using the learning technology back into the data. Collecting data across multiple modalities has been shown to be particularly important for temporal analytics, as it allows researchers to analyze learning phenomena at multiple levels of granularity (Blikstein & Worsley, 2016).

In practice, however, the increased complexity resulting from the additional collection of multimodal data presents unique challenges. Data from different sources are often difficult to integrate. Making sense of all the richness that exists in multimodal data can be highly time-consuming. Imagine trying to understand the detailed sequence of events a student exhibits as he/she engages in productive struggle with a difficult concept and the social interactions surrounding this effort. To fully understand the events that unfold even in this small segment of a student’s educational experience, a researcher may need to watch screen video data and listen to the audio dialogue several times and enter behavioural codes into a separate document. Having to do this for every problem and concept a student experiences over the course of even one class period of learning technology use would be vastly taxing on human time and effort. Yet it is this level of detailed analysis that provides the most interesting and temporally rich insights (Worsley, 2014), in contrast to purely quantitative models so often based solely on coarse-level “correctness” coding.

In this paper, we address this challenge by describing a generalizable approach for combining quantitative and qualitative analyses to yield efficient yet rich sensemaking around intelligent tutoring data. This multi-step approach involves first identifying and flagging “focal points” worthy of more in-depth analysis based on visual (or quantitative) examinations of coarse level (e.g., correctness coding based) learning trajectories. This initial process is adapted from prior work by Stamper & Koedinger (2011). These focal points are then used to define a temporal range of activity to which a more fine-grained, context-rich analysis is then applied. We also describe open-source tools that we are developing to facilitate the extraction of multimodal (audio, video) data segments relevant to these focal points once they have been identified. While the vast majority of multimodal learning analytics research focuses on collecting data from in-classroom activities using a variety of sensory channels (Blikstein & Worsley, 2016), our present work focuses on how low-cost, less-invasive methods to use multimodal data to enrich our understanding of student learning processes while engaged in intelligent tutoring systems. We demonstrate a generalized method for expanding upon the typical quantitative modelling applied to intelligent tutoring systems by incorporating quantitative richness from multimodal data streams, while at the same time minimizing the deployment and processing effort typically associated with other work in multimodal learning analytics.

The first-pass analysis within our approach involves investigating learning trajectories that plot students’ first-attempt correctness across time for each distinct problem step or task. When developers have tagged tasks with knowledge components (the skills, concepts, and/or facts required to complete a given task or problem correctly; Koedinger, Corbett, & Perfetti, 2012), it is most meaningful to visualize and investigate learning trajectories aggregated at the knowledge-component level. In an ideal learning environment, we expect monotonically increasing performance across successive opportunities to apply each knowledge component. Thus, identifying points of disruption to this expectation (particularly

sharp drops in success rate) along learning trajectories for any knowledge component can yield focal moments worthy of more detailed examination.

Often, these points of disruption to an otherwise smooth, monotonically increasing learning curve often are indicative of hidden difficulty factors (i.e., an additional knowledge component) not yet accounted for in the concept-to-problem mapping. In addition, they may indicate that the instructional scaffolding leading up to that moment was less effective than desired. In both cases, deeper investigation is warranted to discover actionable insights relating to students' conceptual struggles at those points in their learning experiences. A more detailed analysis of those focal points, as well as the learning processes building up to them, can be done, either by digging deeper into the process data or the additional multimodal streams.

We discuss the application of this approach to data collected from a classroom study in which students engaged in a "Chemistry Virtual Lab" tutoring system (Davenport et al., 2014, 2015). Detailed log data, as well as screen video and audio data, were collected. We used coarse-level analyses of first-attempt correctness learning trajectories to flag two temporal segments of student activity as worthy of detailed analyses (within our multimodal data streams). We then show that a temporally detailed, multimodal analysis of students' productive struggles within these productive segments led to the identification of informative and actionable insights. These insights had important implications for instruction (both for teachers and the learning technology itself) and led to improvements to the knowledge component mappings used to drive the curriculum and model student learning.

The approach illustrated in the present study is simply one detailed example of how a multi-step approach moving from coarse-level to fine-grain data levels can be an effective way to handle large volumes of rich data at multiple levels of granularity while producing interesting and actionable insights about student learning as it unfolds across time. In the paper, we discuss several suggestions for ways the approach can be adapted to analyze other datasets even larger in volume than the one we present here.

2. Methods

2.1. Chemistry Virtual Lab Tutor

ChemVLab+ (chemvlab.org) provides a set of high school chemistry activities designed to build conceptual understanding and inquiry (Davenport et al., 2014, 2015). Conceptual understanding in chemistry requires students to connect quantitative calculations, chemical processes at the microscopic level (e.g., atoms and molecules), and outcomes at the macroscopic level (e.g., concentrations, colour, temperature). ChemVLab+ is designed to help students connect procedural knowledge of mathematical formalisms with authentic chemistry learning by allowing them to design and carry out experiments. In each activity module, students work through a series of tasks to solve an authentic problem and receive immediate, individualized tutoring. As students work, teachers are able to track student progress throughout the activity and attend to students that may be lagging behind. Upon completion of the activities, students receive a report of their proficiency on targeted concepts and skills, and teachers can view summary reports that show areas of mastery or difficulty for their students. In the current study, students completed four modules: 1) PowderAde: Using Sports Drinks to Explore Concentration and Dilution, 2) The Factory: Using a City Water System to Explore Dilution, 3) Gravimetric Analysis, and 4) Bioremediation of Oil Spills.

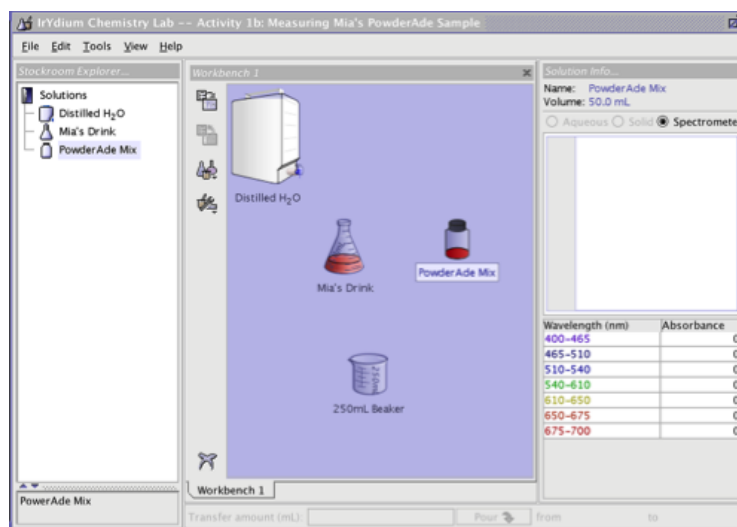


Figure 1. Example interface for the experimentation portions of the ChemVLab+ tutor.

There are a variety of types of interfaces across the four modules, but students spend a significant portion of their time working in open-ended activities such as setting up experiments in a “virtual” laboratory environment (e.g., Figure 1) and making observations.

2.2. Classroom Study

Participants were 59 students at a high school in the greater Pittsburgh area enrolled in honours chemistry classes. They participated in four Stoichiometry modules of the ChemVLab+ educational tutor. They completed these modules across four 50-minute class periods spread over the course of three weeks.

Before students engaged with ChemVLab+, they each completed a paper pre-test assessment. After the four class periods devoted to using the tutoring system, students completed a paper post-test assessment. The vast majority of students completed all four Stoichiometry modules prior to completing the post-test assessment.

Using Camtasia, we collected screen video and audio captures for 47 consenting students during the second and fourth class periods of the study. These multimodal data streams covered the second and fourth ChemVLab+ activity modules (The Factory: Using a City Water System to Explore Dilution and Bioremediation of Oil Spills) for the majority of students. As a result, we focused our present analyses on the data from these two modules.

2.3. Multimodal Data Processing

The captured videos were exported from Camtasia such that each session resulted in one mpeg4 screen video for each unique student and session of ChemVLab+ use. A total of 94 video files, spanning all consenting students for two different classroom sessions, were exported from Camtasia.

Integrating additional information about students’ experience interacting with the system and the learning environment would provide a richer, more complete picture of student learning, especially the process of learning as it unfolds across time. It can, however, be challenging to collect, integrate, and analyze this kind of multimodal data. In particular, data collected from different sources and at different temporal grain sizes (e.g., behaviours quantified at the level of class periods, sections, problems, or sub-problem clicks) can be difficult to integrate. Coding and processing multimodal data streams, like audio and video recordings, for actionable insights can be extremely time-intensive.

To address some of these challenges for researchers, we developed a set of tools called STREAMS (Structured TRansactional Event Analysis of Multimodal Streams) to facilitate the integration of log data and multimodal data streams to produce analyses that uniquely leverage the strengths of both (Liu, Davenport, & Stamper, 2016; <http://github.com/ranolabar/STREAMS>). STREAMS supports 1) easy temporal alignment of software-logged usage data to any number of additional data streams, 2) log data query based extraction of video segments, and 3) visualization of the synchronized streams for exploratory analyses.

The first component of STREAMS accomplishes temporal alignment, where different multimodal streams of data (video, audio, etc.) can be temporally synced with log data and, consequently, to each other. It uses the relative times between log data events, combined with the temporal offset between the logged data and the beginning of each media stream, to do this. If the temporal offset is not automatically recorded during data collection, then minimal human input is required to provide the time within each media stream at which the first software-logged event occurs. The output of temporal alignment is a data frame that contains the original log data, but with three additional columns per synced media stream: the corresponding media stream’s filename, the start time of the event within that stream, and the end time of the event within that stream. Once the data streams of interest are temporally aligned, one can either extract segments of audio/video pertaining to target events, or visualize the synchronized streams of data for exploratory analyses and to create additional annotations.

In the targeted event extraction component of the tool, the user can query any value of any column from the software-logged data (e.g., all problem steps tagged with skill X) or any combination of column values (e.g., all problem steps tagged with skill X on which the student made an incorrect first attempt). STREAMS will then produce a folder of extracted video segments that correspond specifically to the events specified in that query.

Finally, the tool can generate a plugin to DataVyu (DataVyu Team, 2014), a freeware tool that allows different data streams (including audio, video, physiology, eye tracking, motion tracking, and text annotation) to be synced in a manner that allows for easy exploratory analyses and additional annotations across streams. As it exists, DataVyu requires users to manually enter annotations. The STREAMS plugin can, however, extract data from any number of desired log data columns and automatically annotate the multimodal streams with this information within DataVyu. The result is a temporally synchronized collection of both text and multimodal data streams within an interface where additional annotations are easy to create.

Using the temporal alignment function of the STREAMS tool, we integrated all video files with their corresponding segments of the log data. Aligning these video files, totaling over 80 hours of student activity, to the events in the usage log data using STREAMS required just under one hour of human input.

2.4. Analyses

When the content creators (usually domain experts) of an intelligent tutoring system have tagged tasks with knowledge components (the skills, concepts, and/or facts required to complete a given task or problem correctly), it can be useful to visualize and investigate learning trajectories aggregated at the knowledge component level. For example, the knowledge components covered by the ChemVLab+ activities in the present study include unit conversions, concentration, balancing chemical reactions, using stoichiometry to find unknowns, molar mass, investigation and experimentation, and significant figures.

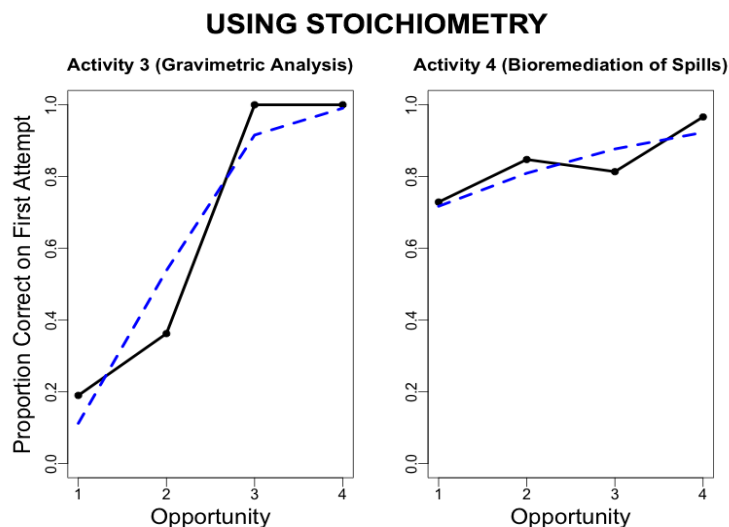


Figure 2. Example of a knowledge component (Using Stoichiometry) for which students show a good learning trajectory within each activity. The solid black line shows actual student first-attempt performance across opportunities to apply the knowledge component to a task. The dotted blue line shows the Additive Factors Model's prediction of each learning trajectory. There was a small drop in success rate for the knowledge component from the last opportunity in activity 3 to the first opportunity in activity 4. This may have been driven by the context change and/or some forgetting between the two sessions (there was about a week between activities 3 and 4)

If underlying knowledge components are well mapped to the tasks or problems that students engage in during learning, then in an ideal learning situation, we expect monotonically increasing performance across successive opportunities to apply each knowledge component. Figure 2 shows an example of reasonably good learning trajectories for what appears to be a well-defined knowledge component from the ChemVLab+ activities.

Sharp drops in performance along a learning trajectory for a given knowledge component often serve as a good indicator of a “focal” point for more fine-grained analysis. They may reveal a hidden difficulty factor (e.g., an additional knowledge component) not yet accounted for in the concept-to-problem mapping (Stamper & Koedinger, 2011). Or, they may indicate that the instruction leading up to that moment was less effective than desired. In both cases, deeper investigation is warranted to understand, at a more detailed level, the cause of such learning trajectory anomalies.

In the first pass of our analysis, we examined the learning trajectories (initially aggregated across all students) for each knowledge component to identify anomalous points. We identified two knowledge components in particular that exhibited one or more sharp “drops” amidst an otherwise smooth and increasing learning trajectory.

Since the ChemVLab+ tutor had relatively few (seven) knowledge components in the initial knowledge component mapping to tasks, we were able to manually examine aggregate learning trajectories for each knowledge component. In the case of a large number of knowledge components, it may be helpful to fit a statistical model to the data and use the parameter estimates to help identify candidate learning trajectories to examine further.

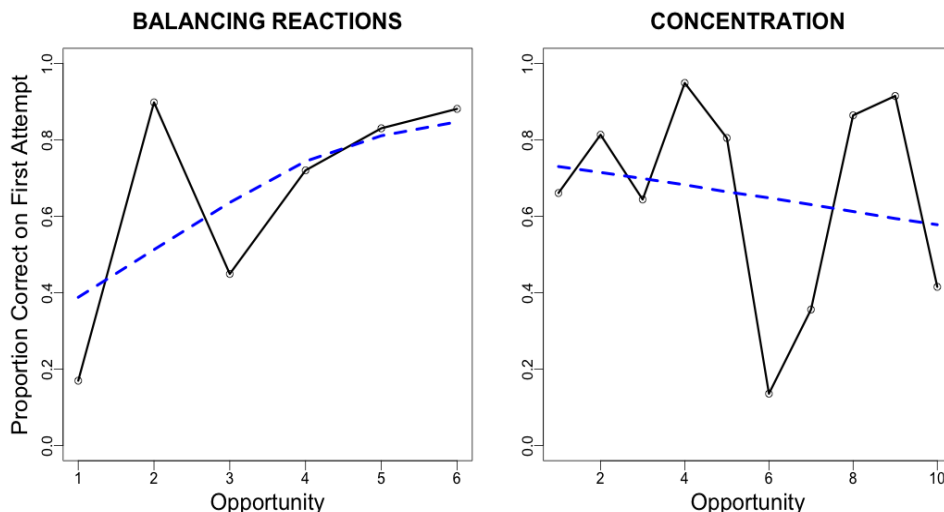


Figure 3. The two knowledge components (Balancing Reactions, Concentration) we identified in which we observed anomalous “drops” along the learning trajectories. The solid black lines show the students’ actual first-attempt performance across opportunities to apply the knowledge component to a task, whereas the dotted blue line shows the Additive Factors Model’s prediction of each learning trajectory.

In Figures 2 and 3, we plotted the model-fitted learning trajectories based on fitting the Additive Factors Model (Cen, Koedinger, & Junker, 2006) to the data. The Additive Factors Model is a logistic regression model that extends item response theory by incorporating a growth or learning term, making it particularly well-suited for modelling temporal data.

$$\ln\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \theta_i + \sum_{k \in KCS} Q_{jk}(\beta_k + \gamma_k T_{ik}) \tag{1}$$

This statistical model (Equation 1) gives the probability p_{ij} that a student i will get a problem step j correct based on the student’s baseline ability (θ_i), the baseline easiness (β_k) of the required knowledge components on that problem step (Q_{jk}), and the improvement (γ_k) in each required knowledge component with each additional practice opportunity. This slope, or “learning rate,” parameter is multiplied by the number of practice opportunities (T_{ik}) the student already had on it. A knowledge component with a good learning trajectory will typically have a slope parameter estimate that is positive.

Thus, one quantitative way to identify knowledge components with problematic learning trajectories is to look for those with negative slope estimates (as for the Concentration knowledge component shown in Figure 3). Another way to identify knowledge components with anomalous points along the learning trajectory is to look for high average residual absolute values. Absolute values of residual values (differences between predicted and actual values) are a way to quantify divergence from the model-predicted learning trajectories. Both the Balancing Reactions and Concentration knowledge components, which had the noisiest learning trajectories, show substantially higher averages of residual absolute values compared to the Using Stoichiometry, which had relatively smooth learning trajectories within each activity.

Table 1. The Additive Factors Model’s slope estimates and average residual values for each of the knowledge components plotted in Figures 2 and 3. Negative slope values and higher average residuals are quantitative indicators of a “noisy” or non-monotonically improving learning trajectory. Absolute values are taken for the residual values to quantify divergence from the model-predicted learning trajectories.

| Knowledge Component | AFM estimate for Knowledge Component Slope | Average of Residual Absolute Values |
|---------------------------|--|-------------------------------------|
| USING STOICH (Activity 3) | 2.263 | 0.757 |
| USING STOICH (Activity 4) | 0.523 | 0.617 |
| BALANCING REACTIONS | 0.260 | 1.013 |
| CONCENTRATION | -0.078 | 1.079 |

For each knowledge component, we identified the points at which the success rate dropped most extremely (opportunity 3 for Balancing Reactions, and opportunities 6–7 for Concentration). These served as focal points around which we conducted deeper investigations looking at a more detailed grain size of data that contained more temporal and contextual richness. We used our Camtasia-recorded screen video and audio data to conduct this deeper level of analysis.

Using the video extraction functionality of the STREAMS tool, we extracted video segments specifically showing those moments of activities plus the previous two opportunities the student had with each knowledge component. Examining immediately preceding opportunities provides a way to discover either contrasting content (allowing for the discovery of hidden difficulty factors) and/or insufficient instructional scaffolding leading up to the problem with which the student is struggling.

Two researchers, one of whom was a chemistry domain expert (and involved in the creation of ChemVLab+) watched the video segments from both knowledge components' focal segments and agreed upon a set of common misconceptions and other relevant behaviours for each. Two video coders then independently coded the video segments based on the presence of these misconceptions and other behaviours of interest. Any discrepancies in the video codes were then debated and an agreement reached between coders. More details on the video coding protocol are described in the Results section.

3. Results

3.1. Balancing Reactions

Problems requiring the Balancing Reactions knowledge component involved students examining a visualization of a chemical reaction and having to produce a balanced equation describing the reaction. To set up the equations correctly, students needed to consider which molecules participated in the reaction. Since the uncharacteristic drop in success rate along their learning trajectories occurred at opportunity 3, we queried the STREAMS tool's video extraction functionality for segments of video pertaining to the 1st, 2nd, and 3rd opportunities students had to apply this knowledge component. Figure 4 shows the problem interfaces for these three opportunities, with the correct solutions described in the figure caption.

3.1.1. Understanding the source of the learning trajectory anomaly

The most common incorrect strategy students exhibited was mistaking the chemical equation as describing the state of the system (e.g., how many molecules are present in the reactant and the product containers) with the process of a reaction (e.g., the rules by which molecules combine). The videos revealed a variety of indicators of this misconception. Many students approached the problem by using the diagrams very literally. For example, in opportunity 1, they would count the A and B molecules in the reactant containers and use those as coefficients for A and B on the reactant side (on the left), and they would count the AB₂ molecules in the product container and use that as the coefficient for AB₂ on the product side (on the right). This process was evident in the video clips as the cursor was visibly moving over the diagram on the screen, and students could be heard quietly counting. Based on our video analysis, 27 (out of 33 total students who got this step incorrect on the first attempt and for whom we had screen video data) exhibited evidence of this misconception on opportunity 1, and 20 (out of 27 total students who were incorrect on the first attempt and had screen video data) exhibited evidence of this misconception on opportunity 3.

It was apparent from our video analysis that students performed much better on opportunity 2, without necessarily needing to understand the relevant concept, for two reasons. First, there were no reactants left over following the reaction, and thus no opportunities for confusion with respect to whether leftover reactants are part of the product side of the equation (many students made the mistake of including leftover reactants in product side as part of their "literal" diagram interpretation). Second, if students applied the literal diagram misconception, they would have typed "10A + 10B => 10AB₂." However, the interface does not allow two-digit coefficients. The inability to respond with a two-digit number indicated to some students that the purely visual strategy was not correct, leading them to reduce the equation before submitting their first attempt. Other students didn't understand why the interface would not accept the 0 but would appear to give up and simply submit 1A + 1B = 1AB₂ which, incidentally, was the correct answer.

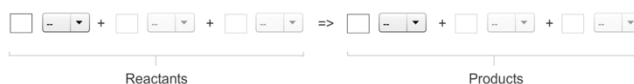
The video analysis reveals that the "drop" in success rate in the Balancing Reactions learning trajectory was primarily driven by the anomalous "spike" in success rate at opportunity 2. This finding suggests that redesigning opportunity 2 to include both leftover reactants and a different number of AB₂ molecules in the product part of the diagram (e.g., 9 rather than 10) would allow researchers to better differentiate between levels of student understanding.

Chem VLab : Stoichiometry Activity 4 : Screen 2 of 20 - chemE_ABZ

(A) Opportunity 1

The diagram to the right shows a reaction between molecules A and B. Use the spaces below to describe this reaction.

Although there are spaces for three reactants and three products, you may not need to use them all.

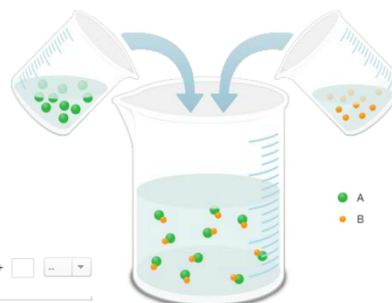
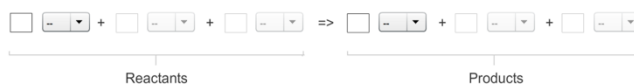


Chem VLab : Stoichiometry Activity 4 : Screen 4 of 20 - chemE_AB

(B) Opportunity 2

The diagram to the right shows a different reaction between molecules A and B. Use the spaces below to describe this reaction.

Although there are spaces for three reactants and three products, you may not need to use them all.



Chem VLab : Stoichiometry Activity 4 : Screen 6 of 20 - chemE_A2C

(C) Opportunity 3

The diagram to the right shows a different reaction. Use the spaces below to describe this reaction.

Although there are spaces for three reactants and three products, you may not need to use them all.

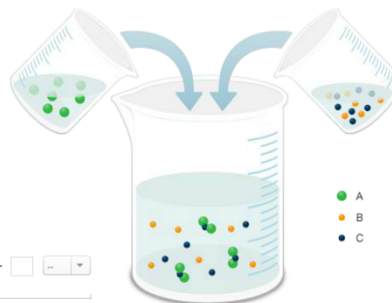
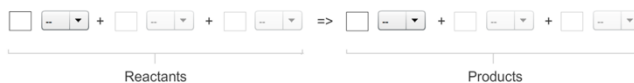


Figure 4. Problem interfaces for the first three opportunities students had to apply the Balancing Reactions knowledge component. These are opportunities leading up to the anomalous spike and drop in students' learning trajectories for this knowledge component. We extracted and analyzed video data of students working on these opportunities. The correct solution to the opportunity 1 problem is $1A + 2B \Rightarrow 1AB_2$. The correct solution to the opportunity 2 problem is $1A + 1B \Rightarrow 1AB$. The correct solution to the opportunity 3 problem is $2A + 1C \Rightarrow 1A_2C$.

3.1.2. Discovering key factors determining students' conceptual evolution over time

Since student performance on opportunities 1 and 3 were the only valid assessments of having a full understanding of how to balance chemical equations based on a diagram of a reaction, we took this opportunity to further investigate students' conceptual evolution across these timepoints. The vast majority (85%) of students got opportunity 1 incorrect on their first attempt, so we wondered whether student behaviours in help-seeking and learning from their mistakes on opportunity 1 had an effect on their first-attempt success on opportunity 3.

To explore this question, we plotted separate learning trajectories for three subgroups of students: those who got opportunity 1 correct on first attempt, those who got opportunity 1 incorrect but demonstrated concrete evidence of understanding the correct solution before moving on to opportunity 2 (based on behavioural coding of video activity), and those who got opportunity 1 incorrect and did not demonstrate concrete evidence of mastering the concept. These learning trajectories are plotted in Figure 5.

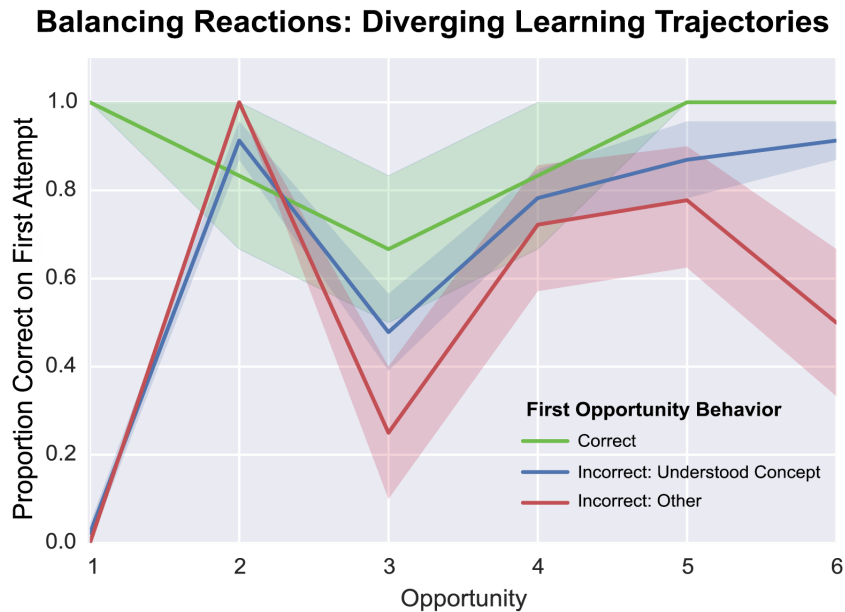


Figure 5. Problem interfaces for the first three opportunities students had to apply the Balancing Reactions knowledge component. These are opportunities leading up to the anomalous spike and drop in students’ learning trajectories for this knowledge component.

Students were coded as demonstrating concrete evidence of understanding the correct solution following an initial incorrect attempt if they 1) were able to reach the correct answer through a logical progression of hint-viewing and updating their attempted answers, without reaching a bottom-out hint (one that essentially provides the answer), or 2) sought help from the teacher or experimenter and verbally demonstrated eventual understanding of the correct solution.

Students who did not satisfy one of these two criteria were coded as having not demonstrated concrete evidence of mastering the concept. The majority of these students either required a bottom-out hint to reach the correct solution or seemed to reach it by trying a series of unprincipled guesses.

As is evident in Figure 5, not only does students’ initial success on their first opportunity strongly relate to their later learning trajectory, but different behaviours in help-seeking and learning among those who made initial mistakes on opportunity 1 affect the learning trajectories as well. Among just those students who were initially incorrect on opportunity 1, we found that their behaviour on this opportunity significantly predicted the learning gains based on pre-test–post-test assessments. In a linear regression model predicting scores on post-test items related to the balanced reactions knowledge component, while controlling for pre-test scores, the student’s first opportunity behaviour had a significant effect ($\beta = 0.24$, $p = 0.0003$; $R^2 = 0.32$).

The cognitive explanation for the divergent learning trajectories is that students who fail to learn from their mistakes on opportunity 1 are going into future opportunities with lower prior knowledge (Hoz, Bowman, & Kozminsky, 2011), and this effect may multiply. A metacognitive explanation (Butler & Winne, 1995) might be that students with better help-seeking behaviours (leading them to better understand their mistakes on opportunity 1) may be better learners in general. The precise attribution of the causes of these divergent learning trajectories (and snapshot assessments of learning) is beyond the scope of the data we collected in the present study.

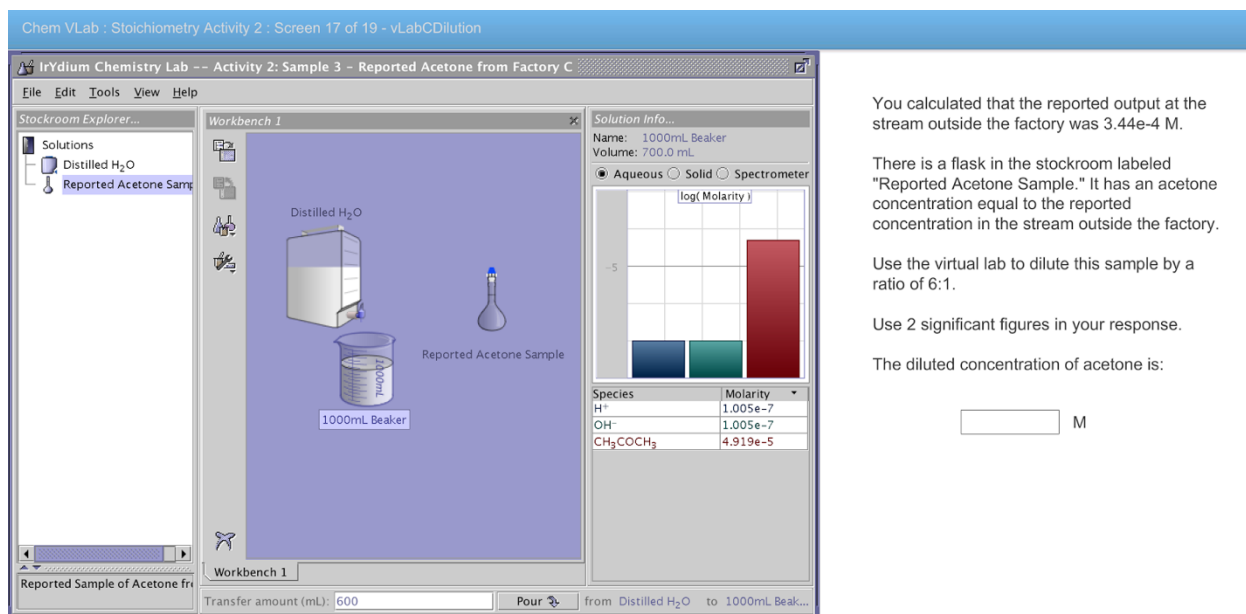
Regardless, there are clear implications of these results for instructional intervention. Opportunity 1 is a critical point whose instruction can potentially influence students’ future learning trajectories. Instruction at this step could be revised to help learners learn from their mistakes. For example, hint messages could be modified to better target dispelling the common misconceptions — like the “literal” diagram interpretation. Bottom-out hints could also be accompanied by a more comprehensive explanation of the concepts.

Analyses of selected video segments of opportunities 1–3 of the Balancing Reactions knowledge component were demanding of human time, but much less so than it would have been to do such a thorough behavioural coding of all 80 hours of video data. Ultimately, the video analysis of this small portion of our data yielded rich insights that were highly interpretable, actionable, and predictive of traditional assessment outcomes (post-test scores). Using a first-pass aggregate learning trajectory analysis to identify these focal segments is an efficient way of using rich data at multiple levels and grain sizes effectively, helping to focus and limit time-intensive analyses, while still yielding significant and actionable discoveries.

3.2. Concentration

The Concentration knowledge component represents understanding that the measure of concentration is the amount of substance (e.g., a sports drink powder) in a volume of substrate (e.g., water). Designed to target conceptual understanding of concentration, the knowledge component includes the ability to read, report, and compare concentrations of solutions. We identified the sharpest drop in the Concentration learning trajectory at opportunities 6 and 7, so we queried the STREAMS tool's video extraction functionality for segments of video pertaining to opportunities 4–7 that students that had to apply this knowledge component.

Behavioural analysis of the video data revealed that students were particularly confused by problems in which a dilution ratio or “factor” is involved (e.g., Figure 6). Many students regarded concentration as a ratio between two volumes rather than a ratio of the amount of substance in a volume. Students exhibited this confusion in responding to prompts such as “Create a 1:2 dilution of the reported sample” or “Add water to the sample until the concentration is diluted by a factor of 2.” The correct solution to these prompts requires students to know that the amount of substance (e.g., the powder) takes up negligible volume, so to dilute the powder by a factor of 2, the total amount of water needs to be doubled. Instead, the majority of students demonstrated shallow knowledge by responding to prompts like these by adding two parts water to one part solution rather than adding one part water to one part solution, which halves the concentration. In another example prompt — “Dilute this sample by a ratio of 6:1” (Figure 2) — students tended to add 6 parts water to 1 part solution, rather than doing the correct action of adding 5 parts of water to 1 part solution.



Chem VLab : Stoichiometry Activity 2 : Screen 17 of 19 - vLabCDilution

IrYdium Chemistry Lab -- Activity 2: Sample 3 - Reported Acetone from Factory C

File Edit Tools View Help

Stockroom Explorer...

- Solutions
 - Distilled H₂O
 - Reported Acetone Sample

Workbench 1

Distilled H₂O

Reported Acetone Sample

1000mL Beaker

Solution Info...

Name: 1000mL Beaker
Volume: 700.0 mL

Aqueous Solid Spectrometer

log(Molarity)

Species Molarity

| | |
|-----------------------------------|----------|
| H ⁺ | 1.005e-7 |
| OH ⁻ | 1.005e-7 |
| CH ₃ COCH ₃ | 4.919e-5 |

Transfer amount (mL): 600 Pour from Distilled H₂O to 1000mL Beak...

You calculated that the reported output at the stream outside the factory was 3.44e-4 M.

There is a flask in the stockroom labeled "Reported Acetone Sample." It has an acetone concentration equal to the reported concentration in the stream outside the factory.

Use the virtual lab to dilute this sample by a ratio of 6:1.

Use 2 significant figures in your response.

The diluted concentration of acetone is:

M

Figure 6. Example screenshot from the screen video of a problem tagged with the concentration and dilution knowledge component. This student incorrectly added six parts water to one part solution. The correct approach was to add five parts water to one part solution.

Interestingly, we observed this confusion on opportunities 6–7 (i.e., the big drop in success rate) but not opportunities 4–5, which did not require a conceptual understanding of dilution factors. These observations are highly indicative that there is a hidden difficulty factor (in this case, understanding dilution ratios) on certain problems tagged with the knowledge component but not others. Following methods described in Stamper and Koedinger (2011), we split the Concentration knowledge component into two distinct knowledge components: one where the problem step required a conceptual understanding of dilution ratios (Concentration-Dilution) and cases where it did not (Concentration-Only). The learning trajectories for the resulting two knowledge components are shown in Figure 7. The trajectories are significantly smoother and show more of an upwards trend than the original Concentration knowledge component's learning trajectory (as shown in Figure 3).

Furthermore, a comparison between the Additive Factors Model fit with the originally defined knowledge components and the same model fit with the revised knowledge components showed a quantitative model fit improvement based on AIC (Akaike Information Criterion; Bozdogan, 1987), which was 1861.5 for the original and 1704.4 for the revised knowledge components. Lower AIC values are indicative of a better model fit.

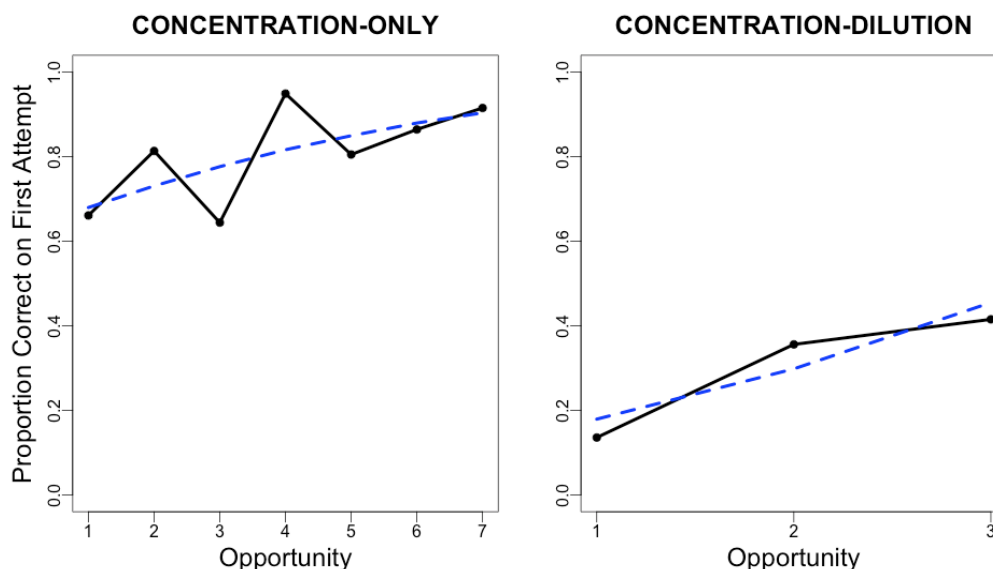


Figure 7. Revised learning trajectories after separating out the original Concentration knowledge component into two separate knowledge components, one reflecting problems that did not require an understanding of dilution ratios (Concentration-Only) and one reflecting problems that did (Concentration-Dilution). These learning trajectories are smoother, and are generally increasing in success rate, compared to the original learning trajectory (depicted in Figure 3).

This discovered revision to the knowledge component-to-problem mapping has several implications for revising instruction. Within the ChemVLab+ technology, recommended points of redesign include adding more practice problems for the newly split Concentration-Dilution knowledge component, which currently has relatively few opportunities for practice and overall low success rates, and adding more hints and instructional scaffolding for the Concentration-Dilution problems. For teachers, this insight implies more instructional attention to be spent on conveying the concept of dilution ratios, with particular emphasis on dispelling students' common misconceptions.

The video analyses greatly facilitated recognizing the specific conceptual struggles that students were experiencing towards this actionable insight that also helped create a more accurate model of student learning. For example, simply analyzing the incorrect answer that students gave for a problem like that shown in Figure 6 would have given no clues as to the nature of their difficulty (which surrounded dilution ratios). We had to watch their virtual lab activities to recognize that students were mistakenly adding 6 parts water to 1 part solution. From the log data, the clickstream data would have had to be reconstituted into a detailed replay for researchers to make high-level sense of what students were doing wrong. Constructing such a replay is not trivial. At the same time, applying human understanding to such video and/or replay data is time consuming.

Starting with the high-level learning trajectory analysis combined with the STREAMS video extraction tool, we are able to quickly “zoom in” to each segment of experimentation activity leading up to students entering an incorrect final answer into the text box. This allowed us to quickly identify the specific misconception, evident in students’ experimentation patterns, that led to difficulties. As a result, recommendations can be made to ChemVLab+ developers and chemistry teachers to modify instruction to confront these misconceptions head on.

4. Discussion and Conclusions

We have illustrated two different routes by which the approach of first identifying “focal points” using knowledge component level trajectories, and then drilling down to the details in the temporal range leading up to those focal points, efficiently led to interpretable and actionable insights while making use of the richness of the available data.

The analysis and coding of video data we did on the Balancing Reactions knowledge component was an example of how a student-centred analysis can benefit from such an approach. Our results from this analysis showed how students’ early experiences struggling with a novel concept could significantly affect both their entire learning trajectories within an activity and pre-test–post-test measurements of learning gains related to that concept. It provided evidence of an important “moment” for early instructional intervention.

The analysis we conducted on the Concentration knowledge component was an example of how a knowledge component-centred analysis can benefit from this multi-step approach as well. Our results led to a modification in the knowledge component assignment to problem steps within a ChemVLab+ activity as well as instructional implications for promoting better learning of a previously hidden conceptual difficulty.

Although first-attempt correctness-based analyses are a limited way of analyzing rich learning process data when done in isolation, applying them as a first-pass to visualize learning trajectories allows researchers to efficiently identify focal temporal segments of likely significance. Traditional methods of analyzing video and audio data can provide a very rich picture of students’ learning activities and processes, but they can be extremely time-consuming and high in human effort. We’ve shown here that the combination of both log data and multimodal data can lead us quickly and efficiently to find temporal segments of interest (using log data) and we can use the richness present in multimodal data to fully understand these segments of interest. The STREAMS tool’s video extraction functionality greatly facilitated the integration of multiple data streams within this approach. Furthermore, this approach helps researchers link quantitative and qualitative analysis more seamlessly and creates fluidity in moving between the two.

Even though we worked with a dataset here that had a relatively small number of initial knowledge components (and thus could manually assess the learning trajectory visualizations), we laid out clear quantitative alternatives that can be applied efficiently to discover focal points of analysis within much larger datasets. The present work serves as illustrative examples of how this generalizable approach can help handle large volumes of rich data at multiple levels of granularity, while producing interesting and actionable insights about student learning as it unfolds across time.

We propose that this analytic approach could be useful, more broadly, to any branch of learning analytics research in which data from multiple sources or modalities must be integrated for analysis. This might include the analysis of multimodal data collected from open-ended, project-based environments (Blikstein, 2013; Blikstein & Worsley, 2016) or the use of multimodal data to code students’ affective states and metacognitive behaviours (D’Mello & Graesser, 2010; Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014) during only high-impact segments of learning. For example, extracting and coding video during segments identified by automated, log-data based affect detectors can provide opportunities to validate these detector algorithms, as well as to provide additional training data for them.

In future work, we aim to incorporate the STREAMS tool into LearnSphere/DataShop (www.learnsphere.org; Koedinger et al., 2010), an educational data and analytic tool sharing infrastructure. LearnSphere already offers the ability to view learning trajectories via a user-friendly web interface and to interactively click on points along learning trajectories for further investigation. Incorporating STREAMS functionalities would allow users to view relevant segments of audio, video, or other data streams (e.g., eye-tracking, sensors, dialogue transcriptions) linked to those trajectory points. This would not only be useful to non-programmer researchers but also to instructors, who could view segments of detailed and multimodal data of their students working during specific problems, concepts, or temporal ranges of interest. They could use these to quickly understand common student difficulties and address them in their instruction. As classrooms shift from worksheet-based to increasingly technology-based practice problems and homework, a tool like STREAMS could also provide an easy way for teachers to extract and showcase worked examples, both correct and incorrect, from actual students.

Acknowledgements

We thank Jacklyn Powers for help in collecting the data used in our experiments here.

Declaration of conflicting interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was supported by the National Science Foundation (Grants #DRL-1418072, PI Davenport, and #DRL-1418181, PI Stamper), and the Institute of Education Sciences, U.S. Department of Education, through Grants R305A100069 and R305A170049 to WestEd. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF or IES.

References

- Barbera, E., Gros, B., & Kirschner, P. A. (2015). Paradox of time in research on educational technology. *Time & Society*, 24(1), 96–108. <http://dx.doi.org/10.1177/0961463X14522178>
- Blikstein, P. (2013). Multimodal learning analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (LAK '13)*, 8–12 April 2013, Leuven, Belgium (pp. 102–106). New York: ACM. <http://dx.doi.org/10.1145/2460296.2460316>
- Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>
- Booth, J. L., Newton, K. J., & Twiss-Garrity, L. (2014). The impact of fraction magnitude knowledge on algebra performance and learning. *Journal of Experimental Child Psychology*, 118, 110–118. <http://dx.doi.org/10.1016/j.jecp.2013.09.001>
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345–370. <http://dx.doi.org/10.1007/BF02294361>
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281. <http://dx.doi.org/10.3102/00346543065003245>
- Cen, H., Koedinger, K. R., & Junker, B. (2006). Learning factors analysis: A general method for cognitive model evaluation and improvement. In M. Ikeda, K. Ashlay, & T.-W. Chan (Eds.), *Proceedings of the 8th International Conference on Intelligent Tutoring Systems (ITS 2006)*, 26–30 June 2006, Jhongli, Taiwan (pp. 164–175). Berlin: Springer-Verlag. http://dx.doi.org/10.1007/11774303_17
- Datavyu Team. (2014). Datavyu: A Video Coding Tool. Databrary Project, New York University. Retrieved from <http://datavyu.org>.
- Davenport, J. L., Rafferty, A., Karabinos, M., & Yaron, D. (2015). Investigating chemistry learning using virtual lab activities in real classrooms. Paper presented in the symposium Theory Driven Design of Online Learning Systems, Data Structures and Analytic Techniques at the Annual Meeting of the American Educational Research Association (AERA 2015), 16–20 April 2015, Chicago, IL, USA.
- Davenport, J. L., Rafferty, A., Yaron, D., Karabinos, M., & Timms, M. (2014). ChemVLab+: Simulation-based Lab activities to support chemistry learning. Paper presented at the American Educational Research Association Annual Conference (AERA 2014), 3–7 April 2014, Philadelphia, PA, USA.
- D'Mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, 20(2), 147–187. <http://dx.doi.org/10.1007/s11257-010-9074-4>
- Graesser, A. C., Conley, M., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook, vol. 3: Applications to learning and teaching* (pp. 451–473). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/13275-018>
- Hoz, R., Bowman, D., & Kozminsky, E. (2001). The differential effects of prior knowledge on learning: A study of two consecutive courses in earth sciences. *Instructional Science*, 29(3), 187–211. <http://dx.doi.org/10.1023/A:1017528513130>
- 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 C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.), *Handbook of educational data mining* (pp. 43–56). Boca Raton, FL: CRC Press. <https://pslcdatashop.web.cmu.edu/about/pubs/KBCSL-v6.pdf>

- Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36(5), 757–798. <http://dx.doi.org/10.1111/j.1551-6709.2012.01245.x>
- Koedinger, K. R., D’Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Cognitive Science*, 6(4), 333–353. <http://dx.doi.org/10.1002/wcs.1350>
- Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, 21(10), 749–759. <http://dx.doi.org/10.1016/j.tics.2017.06.002>
- Liu, R., Davenport, J., & Stamper, J. (2016). Beyond log files: Using multi-modal data streams towards data-driven KC model improvement. In T. Barnes et al. (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining (EDM 2016)*, 29 June–2 July 2016, Raleigh, NC, USA (pp. 436–441). International Educational Data Mining Society.
- Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, 17(1): 33–59. <http://dx.doi.org/10.1080/10508400701793182>
- Ocuppaugh, J., Baker, R. S. J. d., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology*, 45(3), 487–501. <http://dx.doi.org/10.1111/bjet.12156>
- Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129–184. http://dx.doi.org/10.1207/s1532690xci2202_1
- Stamper, J., & Koedinger, K. R. (2011). Human-machine student model discovery and improvement using data. *Proceedings of the 15th International Conference on Artificial Intelligence in Education (AIED ’11)*, 28 June–2 July 2011, Auckland, New Zealand (pp. 353–360). Springer. http://dx.doi.org/10.1007/978-3-642-21869-9_46
- Taber, K. S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice*, 14(2), 156–168. <http://dx.doi.org/10.1039/C3RP00012E>
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265.
- Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge (MLA ’14)*, 12–16 November 2014, Istanbul, Turkey (pp. 1–4). New York: ACM. <http://dx.doi.org/10.1145/2666633.2666634>