Big Data and Learning Analytics: A New Frontier in Science and Engineering Education Research

Abstract

One of the noticeable societal trends caused by the rapid rise of computing power is the availability of big data. From the perspectives of four research projects, this symposium addresses an overall question of what big data and associated learning analytics mean to science education research and science teaching and learning in the classroom. Despite differences in science teaching and learning contexts where these projects are situated, they all face similar challenges in (1) identifying constructs of student cognition to promote in technology-enhanced learning environments, (2) creating capacities to collect meaningful data that can be automatically collected in the environments, (3) analyzing a large amount of learning data produced by students as effectively and meaningfully as possible, and (4) visualizing and using results of analyzed data to inform decisions teachers, students, curriculum developers, and researchers make. Each presenter will address these aspects and discuss related findings and future directions. This array of projects will provide the breadth and depth necessary to introduce big data and learning analytics to the community of science education researchers who are interested in implementing them in their own research.

Lee, H. -S., Xie, C., Nourian, S., Cheney, K. R., Moussavi, R., Gobert, J., Gweon, G., Damelin, D., Finzer, W., Lester, J. C., Wiebe, R. N., Smith, A. (2015). *Big data and learning analytics: A new frontier in science and engineering education research.* Symposium presented at the annual meeting of National Association for Research in Science Teaching. Chicago, IL.

Big Data and Learning Analytics:
A New Frontier in Science and Engineering Education Research

For the last decade, the availability, analysis, and use of big data has created fundamental shifts in the information we use to make decisions in our daily lives ranging from election campaigns to targeted marketing strategies employed in commerce. These shifts have been fueled by the rapid rise of computing power which allows instant, immense, and automatic collection, storage, and analysis of a large amount of data. Big data have forced researchers to reformulate ways in which scientific research is carried out. While these shifts are not yet pervasive in educational research, there is a sign that some researchers are taking notice (Martin & Sherin, 2013). In particular, the field of educational technology in science education is a natural fit to addressing this new opportunity because big data collection and analysis capacities can be easily added to many existing technology-enhanced learning environments. The most important advantage associated with big data is scalability as big data can be collected automatically in an computer-enabled learning environment and analyzed in mass without involving additional time and resources even though the sample size grows significantly. However, the analysis of big data is not straightforward and needs careful investigations before it can be useful to researchers, teachers and students, and curriculum and technology developers.

This symposium addresses how this new opportunity can be conceptualized, operationalized, and materialized in order to improve science education research and science teaching and learning. We address this topic from the perspectives of four research projects. Despite differences in science teaching and learning contexts where these projects are situated, they all face similar challenges in (1) identifying constructs of student cognition to promote in technology-enhanced learning environments, (2) creating capacities to collect meaningful data that can be automatically collected in the environments, (3) analyzing a large amount of learning data produced by students as effectively and meaningfully as possible, and (4) visualizing and using results of analyzed learning data to inform decisions teachers, students, curriculum developers, and researchers make. Each presenter will address these four aspects, present related findings, and discuss future directions. Following presentations, Dr. Janet Kolodner will lead a discussion focusing on the challenges and complexities involved in big data and learning analytics. Then, the audience will have opportunities to interact with presenters as well as the discussant in order to synthesize ideas and studies presented in the symposium.

The four projects are selected to represent a broad spectrum of possibilities with big data and learning anaytics in terms of:

- target audience ranging from elementary to high school students
- complex cognition covering inquiry skills, engineering design, mastery of concepts, and drawing
- analytical approach such as Bayesian Knowledge Tracing, machine-learning, net-working, and topological recognition
- analysis purpose ranging from modeling of student learning over time, improving teaching practices, to visualization of processed information.

This array of research projects will provide the breadth and depth necessary to introduce big data and learning analytics to the community of science education researchers who currently are interested in using them in their own research. The topic of this symposium is critical for shaping the next generation science education researchers who can take advantage of the availability of big data associated with learning processes and outcomes.

Uncovering Engineering Learning with Visual Analytics

Charles Xie and Saeid Nourian at The Concord Consortium

Subject/Problem: The rise of engineering education in K-12 schools (NGSS Lead States, 2013) calls for basic research that can advance our understanding about how students learn engineering. A major research focus is on the process of engineering design (National Research Council, 2012). Because of the open-ended, project-based nature of engineering, students can produce a large quantity of data and artifacts while solving a complex design challenge, making it difficult to discern their learning. Visual analytics is a technique of scientific reasoning that uses *visual interactive interfaces* to optimally combine the computational visualization power of the computer and the pattern recognition power of the brain. This paper will demonstrate how visual analytics can be used to study the learning dynamics of engineering design encoded in the fine-grained data logs of the supporting design software that record all of student actions, artifacts, and articulations. These raw process data are difficult to analyze because of their complex, irregular, and personalized forms. *Visual learning analytics* can provide powerful tools for researchers to see patterns and trends in these student data, from which cognitive and learning theories for engineering education can be tested or derived.

The purpose of this paper is to demonstrate the potential of visual analytics as a methodology for assessing learning in engineering design. Our research is focused on two questions: (1) What science and engineering performance indicators can be computed from the fine-grained data stream logged by the design software? and (2) In what visualization should these indicators be represented so that researchers can rapidly sift large datasets? For example, it is, in theory, through the iterative cycles of design that students learn and apply science and math to optimize their engineering solutions and products (National Research Council, 2012). But how do we measure the degree of iteration for each student? These information cannot be easily obtained from pre/post assessments and can only be reliably extracted from the actual process data.

Analytic Methods: Approximately 200 high school students participated in this research project. Each student was challenged to solve an engineering design problem related to energy efficiency of the built environment using our Energy3D software, a modern computer-aided design (CAD) tool that supports form and function design with accurate, real-time scientific simulations while logging all student actions behind the scenes. In total, the students generated more than 2 GB of structured process data, creating a gold mine of data for this research. The following subsections provide some examples that show how visual analytics can be used to construct cognitive or non-cognitive performance indicators.

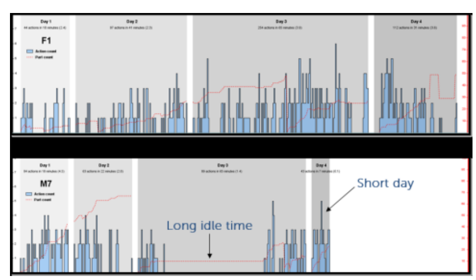

Figure 1. Manhattan plots of students' action time series that represent the level of student activeness during a multi-day engineering design project using Energy3D.

Results: *Action time series measuring student activeness*. As practice is an indispensable part of engineering, student activeness is a critical variable in engineering assessment. The outcomes for students who are deeply engaged and the outcomes for those who are largely disengaged need to be evaluated in separate cohorts. Time series visualization of student action logs provide a simple but reliable way to sort student activeness before carrying out in-depth analyses (Xie, Zhang, Nourian, Pallant, & Hazzard, 2014). For example, by plotting the actions as a function of time as shown in Figure 1, we can clearly see that the first student maintained strong interest throughout the entire project whereas the second student went idle for a long period in the middle of the project.

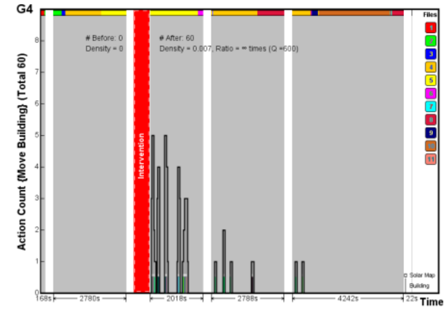

Figure 2. A visualization of a student's response to an intervention (red bar) with a certain type of design action (other bars) during an Energy3D

Response functions measuring intervention outcomes. Measuring the effect of an intervention quantitatively is a central task of educational research. This can be mathematically modeled by using the response function, which describes how a stimulus prompts a design action (Xie, Zhang, Nourian, Pallant, & Bailey, 2014). Our results show that the occurrence of the design actions unrelated to an intervention were, not surprisingly, unaffected by it, whereas the occurrence of the design actions that the intervention targeted revealed a continuum of reactions ranging from no response to strong response. From the temporal patterns of these student responses, persistent effect and temporary effect (with different decay rates) were identified (Figure 2). This result is significant because it demonstrates a technique for determining the effect of formative feedback based on data logs.

Polar plots measuring design space exploration. Engineering design is a creative process. The design space is high-dimensional: A design comprises a number of elements (building blocks) added and revised through a number of actions that set or change their properties. The dynamic change of the volume of the subspace in which a student explores from episode to episode may be characteristic of his/her iterative divergent-convergent design thinking (Dym, Agogino, Eris, Frey, & Leifer, 2005). To visualize the explored design space, we use a polar plot whose axes represent the design dimensions. Dimensional attributes can be drawn on each axis. For example, the number of actions in each dimension can be shown using dots (Figure 3). It becomes immediately clear from the plot how widely the student has explored the design options. As the plot is interactive, clicking each axis will bring up additional information. For instance, a time series graph of the selected actions will be opened to show when the actions occurred and how their occurrences are correlated.

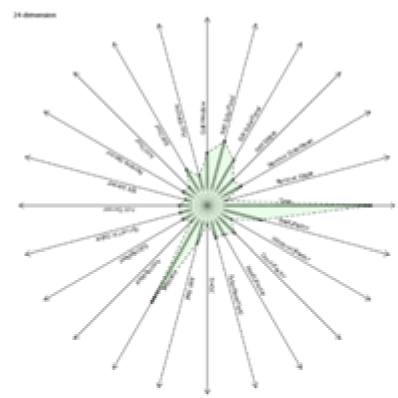

Figure 3: A 24-dimension polar plot

Discussion: As in any other scientific discipline, visualizations are useful tools in educational research. As learner data explode in the era of digital learning (Bienkowski, Feng, & Means, 2012), visual analytics will be important to researchers who work at the intersection between learning analytics and learning sciences (Martin & Sherin, 2013). This methodology will allow researchers to analyze large quantities of data produced by interactive learning environments more efficiently to discover or refine effective methods that foster deep learning in science and engineering practices.

Using Learning Curves to Guide the Teaching of Inquiry Skills

Kyle R. Cheney, Raha Moussavi, and Janice Gobert

Subject/Problem: Science inquiry skills are increasingly emphasized as an important aspect of students' scientific literacy (NGSS Lead States, 2013). However, it can be challenging for students to learn and hone these inquiry skills because of their underlying complexity (e.g. Gobert, Sao Pedro, Baker, Toto, & Montalvo, 2012). Research on the nature of science inquiry skills using learning analytic techniques can help to guide the teaching of these skills in such a way that fosters more effective skill acquisition by the students (Gobert et al, 2012; 2013).

The goal of this paper was to use learning analytics to investigate the extent to which inquiry skills transfer across science domains to better understand the degree to which inquiry skills are domain-specific. Because learning analytic techniques offer a more powerful way of analyzing data at a much more fine-grained level that is not often possible in classroom studies, our findings have clear implications for science instruction.

Thus, the transferability of inquiry skills was investigated by looking at learning curves of students' performances on four virtual labs in an intelligent tutoring system designed to assess inquiry skills. Inq-ITS (Inquiry Intelligent Tutoring System; Gobert, et al, 2012, 2013) engages users in virtual labs, or microworlds, to form hypotheses, experiment by manipulating variables and collecting data, and interpret the results of those experiments within specific science domains (e.g. phase change, Newtonian physics, biology of a cell, or ecosystems). The system uses learning analytics to evaluate log files generated from student interactions within the microworlds as a measure of student performance on 15 different inquiry skills and sub-skills relating to hypothesis generation, experimentation, and data interpretation. Inq-ITS has been shown to be an effective method for assessing inquiry skills in authentic scenarios (Gobert et al, 2012; 2013). For a complete review of the system see Gobert et al., 2013.

Methods: The data used in this analysis were from 155 Central Massachusetts eighth grade students' within Inq-ITS across three physical science topics (phase change, density, and free fall) and one in general inquiry during the 2013-2014 school year in conjunction with the regular classroom curriculum. Each student completed the microworlds in the same order and each session took one class period. The sessions were separated by an average of 56 days.

In order to assess the extent to which inquiry skills transfer across the physical science topics, a learning curve of student performance on the measured inquiry skills was constructed with the PSLC's DataShop (Koedinger et al., 2010). Using DataShop, models and corresponding learning curves of complete, partial, and zero transfer were also constructed. How well the learning curves of each of these simulated models of transfer matched the actual data was examined using Additive Factor Modeling within DataShop.

Each of the three models of transfer was comprised of a different set of knowledge components consisting of a unique combination of the 15 inquiry skills/sub-skills and 4 microworlds. In the complete transfer model, students' performance was grouped only by the 15 general inquiry skills (a 15 knowledge component model), which appear across all four microworlds. In the zero transfer model, all of the inquiry skills were tied to the microworlds in which they occurred, creating a 60 knowledge component model (15 inquiry skills x 4 microworlds). For the partial transfer model, both the inquiry skill and the microworlds were used to create groups additively for a total of 19 knowledge components (15 inquiry skills + 4 microworlds).

Results: Results from the evaluation of the learning curves provides support for the theory that there is at least partial transfer of inquiry skills from one science topic to the other. Using BIC as a measure of model fit, the model of partial transfer outperformed the other two models (partial transfer > complete transfer > zero transfer). Furthermore, analyses of the learning curves of student performance revealed that although there is a slight increase in the error rate as students transition from one microworld (science topic) to another, there is an overall decrease in error across interactions, which is indicative of partial transfer. A learning curve demonstrating no transfer would show a jump back up to the initial error rate after each transition to a new microworld. If there were complete transfer, a relatively even, downward slope would appear across all the sessions.

Discussion: These findings support past research (Chen & Klahr, 1999; Kuhn, Schauble, & Garcia-Mila, 1992) that inquiry skills can transfer from one science topic to another. The use of learning analytics provides an opportunity to look at transfer of inquiry skills at a more fine-grained level than in previous research. These findings also have implications for the way we understand and teach inquiry skills. Since our data suggests that inquiry skills partially transfer across topics, teachers can leverage the benefits of teaching inquiry in one domain when teaching a new domain. This both facilitates teachers' instructional practices and supports students' learning of inquiry. These findings also have implications for Inq-ITS, since students can be reminded by the system about their performance on past inquiry tasks as possible hints.

Though these findings are in line with past research, future analyses will use learning analytics to disentangle the extent to which each of the individual inquiry skills interact with one another. These analyses will also examine if certain inquiry skills transfer across topics better than other inquiry skills (e.g. whether an inquiry skill regarding hypothesis formation is more apt to transfer than an inquiry skill for interpreting data). Furthermore, since students' interactions with the microworlds were separated by an average of 56 days, this may have had an effect on student performance and should be further examined.

Modeling Student Learning of a Mechanical System during Game-like Simulations

Hee-Sun Lee, Gey-Hong Gweon, Dan Damelin, & William Finzer

Subject/Problem: In a technology-enhanced learning environment, students are expected to learn scientific knowledge through interactive tasks configured with games or simulations (National Research Council, 2011). Students' actions with the environment and resulting performances on the interactive tasks can be captured automatically in the background as text-based logs with time stamps. This study addresses an analytic approach called Bayesian Knowledge Tracing (BKT) that can describe how students acquire knowledge over time about a simple mechanical system involving a car on a ramp. The interactive ramp task was developed in a game format where students progress through a number of levels with increased conceptual difficulties. Research questions are (1) How does the BKT quantitatively model student progress in understanding the knowledge about the car-on-ramp system? and (2) What do the BKT results tell about the usefulness of the graphing tool in supporting student learning during the task?

Learning Task: In the ramp task, students were asked to determine a height so that the car could land on a particular location. The ramp task consisted of four performance levels requiring students to apply more and more sophisticated knowledge about the system as follows:

- Level 1: relationship between height and a fixed landing location

- Level 2: relationship between height and moving landing locations
- Level 3: relationship between height and slope of the ramp on a fixed landing location
- Level 4: relationship between height, slope, and friction on a fixed landing location

Each level was comprised of four steps. To help students, a graphing tool was available. The graphing tool did not initially appear to students and thus should be activated by the students. Once activated, the graph would be automatically drawn based on the x- and y-axes students chose. A graph related to height vs. distance would be necessary for Levels 1 and 2; slope vs. height for Level 3; friction vs. height for Level 4. Active manipulation of the graphing tool was necessary for students to succeed. Students' performances were scored automatically on a 0 to 100 scale based on how close the car landed from the specified landing location. If students scored 90 points or higher, then they were allowed to move to the next step within the level. If students finished all four steps within the level, they moved to the first step of the next level.

Methods: The ramp task was implemented in eight classrooms taught by two teachers in two high schools located in the Northeastern part of the US. One teacher taught a 12th grade physics course in a suburban setting while the other teacher taught a 9th grade physics course in an urban setting. The logged data analyzed in this study were 29,110 lines long and belonged to 42 registered student groups who carried out the ramp task in a single class period. The BKT analysis was then applied to each student group's data. The BKT considers a knowledge variable as latent which affects student performance. The BKT estimates the knowledge variable drawn from observed student performances as a function of time. The BKT analysis estimates the knowledge growth using four parameters as follows (Corbett & Anderson, 1995):

- $p(L_1)$: Initial knowledge parameter associated with the probability that the student already knows the target knowledge prior to the task
- p(T): Transition parameter associated with the probability of becoming knowledgeable at a given level
- p(G): Guessing parameter associated with the probability of guessing correctly without the target knowledge (false positive)
- p(S): Slip parameter associated with the probability of making a mistake when in fact the student has the target knowledge (false negative)

These four parameters determine a unique curve for students' knowledge growth within a specified time frame. The larger the guessing parameter, the higher the likelihood of guessing the result correctly without possessing the knowledge. The larger the slip parameter, the higher the likelihood of choosing wrong heights despite having the target knowledge. We segmented each student group's log data by level because each level presented a new piece of physics knowledge for students to learn. We then estimated four BKT parameters within each level of the ramp task. This progression of knowledge mastery across levels was fitted based on Bayesian statistics with the Monte Carlo sampling method. In order to compare the impact of the graphing tool on student learning within the level, we coded a log data segment as "0" if the graphing tool was not actively manipulated in the segment and "1" if the graphing tool was actively manipulated. We then compared each of the four BKT parameters by the graph manipulation variable using mixed effects ANOVA with student group as a random effect and graph manipulation as a fixed effect.

Results: On average, student groups spent an average of 1,663 seconds on the ramp task ranging from 197 seconds to 3,491 seconds (SD = 505 seconds). An average number of log lines analyzed per group was 462 lines with a standard deviation of 179 lines. Among the 42 groups, at the end of one class period, 10% successfully reached the highest level, 31% worked at the Level 4; 38% worked at the Level 3; 19% worked at the Level 2; and 2% worked at the Level 1.

There were 137 segments in this data set. When student groups just started a level and did not conduct an enough number of simulation trials, the BKT algorithms could not estimate L1, T, G, S parameters for that level because parameter fitting did not converge. There were 10 such segments. As a result, we had 127 segments to which BKT parameters were estimated. According to Table 1, mixed effects ANOVA results indicate that, after controlling for group effects, the probability of guessing, p(G), was significantly lower for groups who actively manipulated graphs during a level than those who did not. After controlling for group effects, the probability of slip, p(S), was significantly lower for groups who actively used the graphing tool than those who did not. There were no significant group effects in all four parameter estimate comparisons. No significant interaction effects between group and graph manipulation across four BKT parameter estimate comparisons, indicating that the graph effect did not depend on which group was using graph.

Discussion: Together, these preliminary findings indicate how the embedded graphing tool worked for students: by encouraging students to make informed choices during simulations, rather than relying on random guessing. We believe that the BKT modeling approach can be useful to track student progress in game-like simulations and can facilitate the design-based research by allowing researchers to associate the effect of cognitive tools embedded in an interactive learning environment with learning process indicators captured by the logging data. The sample size of 42 groups limited our ability to investigate how other factors such as students characteristics, behaviors, and task features impacted their learning. Our next research step involves the investigation of whether and how particular patterns of log events are associated with particular types of student discourse (student learning presented in dialogic forms) and behaviors so that we can identify valid and reliable indicators (proxies) of student learning.

Table 1. Mixed effects ANOVA Results on BKT parameter estimates

BKT parameter estimates	Without graph		With graph		Graph effect
	Mean	SD	Mean	SD	F, p
(a) P(Li): Probability of having knowledge before the level	.18	.10	.15	.12	F = .99, p = .32
(b) P(T): Probability of becoming knowledgeable within the level	.56	.18	.52	.18	F = .43, p = .52
(c) P(G): Probability of guessing correctly without knowledge	.37	.29	.26	.23	F = 6.4, p < .05
(d) P(S): Probability of making mistakes with knowledge	.41	.20	.31	.16	F = 11.5, p < .01

Note. All parameters ranged from 0 to 1.

Towards Sketch-based Learning Analytics

James C. Lester, Eric N. Wiebe, & Andrew Smith, North Carolina State University

Subject/Problem: Diagrams and sketching are fundamental to teaching and learning in science education. From primary through post-secondary education, students use drawings and graphical representations to make sense of complex systems and as a tool to organize and communicate their ideas to others. Studies have shown that learning strategies focusing on learner-generated sketches can produce effective learning outcomes, such as improving science text comprehension and student engagement (Rich & Blake, 1994), facilitating the writing process (Moore & Caldwell, 1993), and improving the acquisition of content knowledge (Britton & Wandersee, 1997). Furthermore, spatial ability (facilitated through drawing) has been recognized as a predictor of STEM success even when accounting for mathematical and verbal ability (Wai, Lubinski, & Benbow, 2009).

Unlike the well studied areas of how people learn from writing text, viewing pre-rendered graphics, and reading, relatively little is known about how the generation of scientific drawings affects learning. Van Meter and Garner (2005) posit that students asked to draw a picture engage in three cognitive processes: selecting relevant information, organizing the information to build up an internal verbal model, and constructing an internal nonverbal representation to connect with the verbal representation. Others suggest that drawing can be a meaningful learning activity requiring both essential and generative processing to mentally connect multiple knowledge representations (Schwamborn, Mayer, Thillmann, Leopold, & Leutner, 2010). One of the first steps towards developing sketch-based learning analytics is devising methods for automatically analyzing students' drawings. In this presentation we describe our work on automatically analyzing student-generated science drawings and discuss the rich potential of sketch-based learning analytics.

Methods: We have been exploring sketch analytics in the context of the LEONARDO project (Figure 1), which is designing, developing, and investigating an intelligent virtual science notebook for interactive scientific modeling in elementary science education. Students in Grades 4 and 5 use LEONARDO to create and experiment with interactive models of physical phenomena. With a curricular focus on electricity and magnetism, LEONARDO features a pedagogical agent that provides explanations and advice as students undertake modeling activities. LEONARDO's curriculum is based on that of the Full Option Science System (FOSS) and is aligned with the Next Generation Science Standard goals in elementary school science education. Throughout the inquiry process, students using LEONARDO are invited to create symbolic sketches, including electrical circuits. Given the challenges of machine analysis of freehand sketching, as well as concerns of excessive cognitive demand for elementary students working in such an unstructured space, LEONARDO supports icon-based symbolic drawing tasks. In conjunction with LEONARDO, we have developed SKETCHMINER, a sketch data mining system that automatically analyzes and compares student drawings using topological graphs (Figure 2).

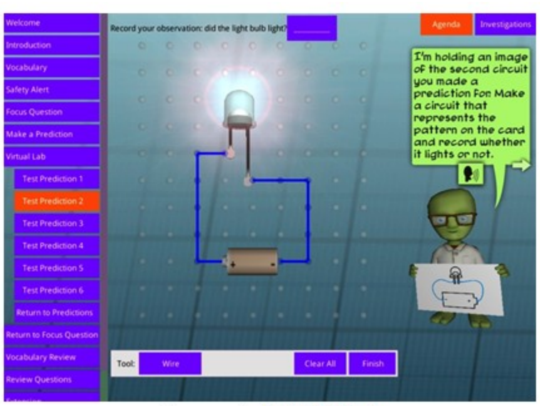

Figure 1. LEONARDO

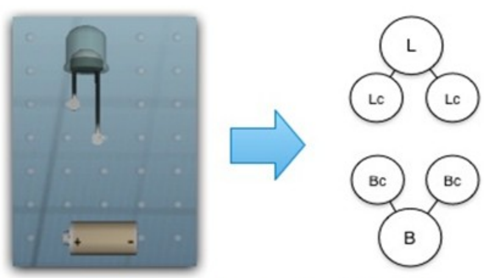

Figure 2. Circuit elements and corresponding topology

For the analyses of SKETCHMINER that will be discussed in the presentation, a corpus of symbolic drawings was collected from fourth grade students interacting with LEONARDO in

North Carolina and California. After data cleaning, drawing activities from 132 students were used for the analysis. Student drawings were scored in comparison to normative models constructed by the research team. To evaluate SKETCHMINER, we clustered student drawings using both an unweighted and a weighted topographical edit distance as the distance metric.

Results: In order to evaluate the clusters, two independent coders from the project's education team developed a rubric and scored the student responses for circuit drawings involving a switch, motor, and battery connected in series. Based on the rubric, the drawings were independently classified into 4 clusters by the two coders ($\kappa = .9$), creating a gold standard clustering to validate our clusters against. SKETCHMINER produced strong alignment with the human classifications, with the weighted edit distance producing better results than unweighted (Table 2). Cluster-based classification accuracies for the highest performing distance metric are shown in Table 3.

Table 2. SKETCHMINER Classification Accuracy

Distance Metric	Accuracy	Precision	Recall
Unweighted	0.73	0.56	0.63
Weighted	0.86	0.74	0.76

Table 3. SKETCHMINER Classification by Class – Weighted Edit Distance

Class	Accuracy	Precision	Recall
1 (Blank)	0.89	0.61	1
2 (No Structure)	0.87	0.66	0.5
3 (Some Structure)	0.86	0.92	0.6
4 (Correct)	0.98	1	0.96

Discussion: The results show promise as a means of automatically assessing learner drawings and suggest several lines of investigation for sketch-based learning analytics. First, while "distance to solution" is a valuable metric, SKETCHMINER's edit distance could also be used to compare errors to each other. Preliminary analysis using this technique has shown promise for identifying common error states that could be used in curriculum redesign or to generate targeted scaffolding for students. Perhaps the most promising area for analysis is investigating the drawing process itself. Topographical representations can be created at any point in the drawing process, allowing for analysis of sequences and patterns in student drawing. Models learned from corpora of learner drawing processes can be used to create more accurate models of learners' conceptual representations, as well as the basis for providing customized scaffolded support scaffolding to a broad range of learner populations.

References

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Washington, DC: Office of Educational Technology, U.S. Department of Education.

Britton, L. & Wandersee, J. (1997). Cutting up text to make moveable, magnetic diagrams: A way of teaching & assessing biological processes. *The American Biology Teacher*, 59, 288-291.

Chen, Z. & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variable strategy. *Child Development, 70*, 1098-1120.

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

Dym, C. L., Agogino, A., Eris, O., Frey, D., & Leifer, L. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education, 94*(1), 103-120.

Gobert, J., Sao Pedro, M., Raziuddin, J., & Baker, R. (2013). From Log Files to Assessment Metrics for Science Inquiry Using Educational Data Mining. *Journal of the Learning Sciences, 22(4)*, 521-563.

Gobert, J., Sao Pedro, M., Baker, R., Toto, E., & Montalvo, O. (2012). Leveraging Educational Data Mining for Real Time Performance Assessment of Scientific Inquiry Skills within Microworlds. Journal of Educational Data Mining, 4(1), 111-143.

Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J. (2010). A Data Repository for the EDM community: The PSLC DataShop. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press.

Kuhn, D., Shauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9,* 285-327.

Martin, T., & Sherin, B. (2013). Learning Analytics and Computational Techniques for Detecting and Evaluating Patterns in Learning: An Introduction to the Special Issue. *Journal of the Learning Sciences, 22*(4), 511-520.

Moore, B. & Caldwell, H. (1993). Drama and drawing for narrative writing in primary grades. *The Journal of Educational Research*, 87, 100-110.

National Research Council, (2011). *Learning science through computer games and simulations.* Washington, DC: The National Academies Press

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington DC: The National Academies.

NGSS Lead States. (2013). *The Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

Rich, R. & Blake, S. (1994). Using pictures to assist in comprehension and recall. *Intervention in School and Clinic*, 29(5), 271-275.

Schwamborn, A., Mayer, R., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology*, 102(4), 872-879.

Van Meter, P. & Garner, J. (2005). The Promise and Practice of Learner-Generated Drawing: Literature Review and Synthesis. Educational Psychology Review, 17(4), 285-325.

Wai, J., Lubinski, D., & Benbow, C. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817-835.

Xie, C., Zhang, Z., Nourian, S., Pallant, A., & Bailey, S. (2014). On the Instructional Sensitivity of CAD Logs. *International Journal of Engineering Education, 30*(4), 760-778.

Xie, C., Zhang, Z., Nourian, S., Pallant, A., & Hazzard, E. (2014). A Time Series Analysis Method for Assessing Engineering Design Processes Using a CAD Tool. *International Journal of Engineering Education, 30*(1), 218-230.