Boise State University
ScholarWorks

University Author Recognition Bibliography: 2018   The Albertsons Library

6-1-2018

A Review of Tools and Techniques for Data-Enabled Formative Assessment

Rob Nyland
Boise State University

A Review of Tools and Techniques for Data-Enabled Formative Assessment

Rob Nyland*
Boise State University
robnyland@gmail.com

Author Biography

Rob Nyland is a PhD candidate in the department of Instructional Psychology & Technology at Brigham Young University. Prior to his PhD he was a tenure track faculty member in Multimedia Design and Production at Lake Washington Institute of Technology in Kirkland, WA, where he was the head of the colleges eLearning council. Previously, he completed a MA in Mass Communications and a BA in Media Arts at BYU. His research focus includes personalized learning platforms, learning analytics, and competency-based education.

Abstract

The purpose of this literature review is to understand the current state of research on tools that collect data for the purpose of formative assessment. We were interested in identifying the types of data collected by these tools, how these data were processed, and how the processed data were presented to the instructor or student for the purpose of formative assessment. We identified two categories of data: machine graded and activity stream data. The data were processed using three methods: unprocessed activity streams, descriptive data analysis, and data mining. Processed data were presented to students through reports and real-time feedback, and to instructors through reports and visual dashboards.

Keywords: data, formative assessment, feedback, dashboards, data mining

Feedback is a critical component of successful teaching and learning. This is powerfully demonstrated in Hattie's (1999) meta-analysis of 196 educational studies in which he found that feedback had an average effect size of .79, nearly double the average effect size of all other interventions combined (.40). He concluded, "The simplest prescription for improving education must be 'dollops of feedback' —providing information [on] how and why the child understands and misunderstands, and what directions the student must take to improve" (Hattie, 1999, p. 11). In the classroom, feedback can be produced through formative assessment which "provide[s] feedback on performance to improve and accelerate learning" (Sadler, 1998, p. 77).

Good formative assessment relies on observation of the student learning process. This observation can be difficult as student work is increasingly done in computer aided environments. While instructors might be able to read the body language or facial expressions of a student in a face-to-face environment to assess understanding, such signs of misunderstanding are not readily present in online environments. We need to harness information collected by these online tools to better enable formative assessment for an instructor.

This need for smarter formative assessment tools is echoed by the U.S. Department of Education (2011) who stated that online learning systems have the potential to "be used formatively to diagnose and modify the conditions of learning and instructional practices while at the same time determining what students have learned for grading and accountability purposes" (pg. xi).

While there has been much research in the last several years regarding the potential for finding patterns in educational data (see Papamitsiou & Economides, 2014), there has not been a systematic review of research regarding data-enabled learning tools designed for formative assessment. In this review, we wanted to understand the current state of the literature regarding the use of technology-provided, data-enabled formative assessment. In particular, we wanted to understand the type of data being collected by these systems, how the data are processed, and how the processed data are presented to either the instructor or the students. By understanding the methods that are currently being employed for the purposes of technology-enabled formative assessment, we hope to identify the best way to enable feedback to students and instructors through these systems.

The literature review is guided by three questions:

1. What types of data are technology-enhanced formative assessment systems capturing?
2. What methods are being used to process the data?
3. Once the data are processed, how are the findings from the data being used for the purposes of feedback to either instructor or student?

Methods

The aim of this systematic literature review was to identify how data-enabled technological systems have been used for the purpose of formative assessment. Because many research fields—including learning analytics, educational data mining, artificial intelligence, and online learning—address topics in technology-enhanced assessment, we began with a broad set of search terms. We searched the terms *learning analytics, data mining, data analysis, assessment, formative evaluation, visualizations, dashboards, intelligent tutoring systems, computer-mediated communication*, and *data analysis* in the following electronic databases: ERIC, Education Full Text (H.W. Wilson), PsycInfo, Computers & Applied Sciences Complete, the ACM Digital Library, and Google Scholar. Further searching was accomplished through backwards referencing of collected studies. The search was not bounded within any specific time period, but due to the types of technologies that we were looking for, articles tended to be from the last 20 years. Only peer-reviewed articles and conference presentations were included in the search.

Inclusion Criteria

After scanning through the initial set of articles, articles were removed according to the following inclusion criteria, matching our review questions:

- The system implemented in the research needed to have an explicit purpose of providing feedback to students or instructors based on data collected in the course.
- The system needed to capture student data digitally (e.g. through computers, mobile devices, or other sensor data) rather than through paper and pencil or Scantrons.
- The articles needed to report on implemented formative assessment systems, rather than theoretical designs.

As a result of our inclusion criteria, several revolutionary built for data analysis purposes, such as HIMATT, AKOVIA, ALA-Reader, and SNAPP were not included. While these tools are helpful in understanding student conceptual knowledge and community structure, they did not directly give formative feedback to students or instructors regarding content knowledge.

After excluding articles that did not meet the criteria, 28 total articles remained for review. Each article was reviewed according to each of the three review questions. Articles were coded for the type of data collected, the method used to process the data, and how the processed data were presented to the instructor or student for the purpose of formative assessment. After categories were developed, 20 percent of the articles were independently examined by another reviewer using the developed codes. Initial agreement between the reviewers was at 80 percent, after which the coding definitions were further clarified and the two reviewers came to complete agreement on the articles in the sample. It was then deemed appropriate that the developed codes could be used on the remaining articles.

Findings

After reviewing the articles, several categories for each of the research questions emerged. In response to the first question, there were two categories of data: machine scored data, and activity stream data. For the second question, there were three data processing methods: activity streams, descriptive data analysis, and data mining. The third question had two sets of categories: methods for presenting data to students and methods for presenting data to instructors. In the student category, there were two presentation methods: reports and dashboards, and real-time feedback. In the instructor category, there were also two presentation methods: reports and visual dashboards. These were broken up into separate categories for the purpose of analysis. The reviewed articles along with their mapping in each of the categories is presented in Figure 1. Note that some articles addressed all aspects of the research questions (e.g., Chen & Chen, 2009), while other articles only addressed one aspect (e.g., Bajzek, Brown, Lovett, & Rule, 2007). In the remainder of this section, we will discuss those aspects of the reviewed studies derived from answering each of the research questions.
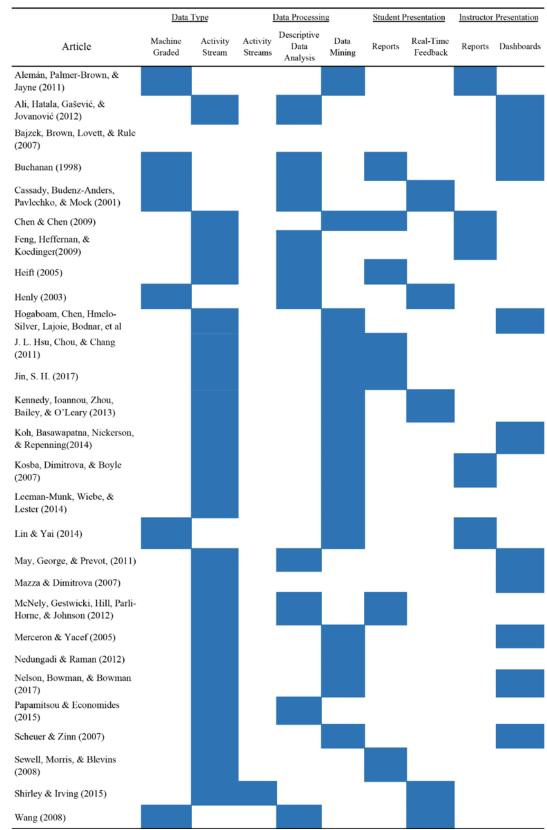
Article	Data Type		Data Processing			Student Presentation		Instructor Presentation	
	Machine Graded	Activity Stream	Activity Streams	Descriptive Data Analysis	Data Mining	Reports	Real-Time Feedback	Reports	Dashboards
Alemán, Palmer-Brown, & Jayne (2011)	X				X			X	
Ali, Hatala, Gašević, & Jovanović (2012)		X		X					X
Bajzek, Brown, Lovett, & Rule (2007)	X			X					X
Buchanan (1998)	X			X		X			X
Cassady, Budenz-Anders, Pavlechko, & Mock (2001)							X		X
Chen & Chen (2009)		X			X		X	X	
Feng, Heffernan, & Koedinger (2009)		X		X	X			X	
Heift (2005)						X	X		
Henly (2003)	X			X			X		
Hogaboam, Chen, Hmelo-Silver, Lajoie, Bodnar, et al		X			X				X
J. L. Hsu, Chou, & Chang (2011)		X			X				
Jin, S. H. (2017)		X			X				
Kennedy, Ioannou, Zhou, Bailey, & O'Leary (2013)		X				X	X		
Koh, Basawapatna, Nickerson, & Repenning (2014)		X			X				X
Kosba, Dimitrova, & Boyle (2007)		X			X			X	X
Leeman-Munk, Wiebe, & Lester (2014)		X			X				
Lin & Yai (2014)	X				X			X	
May, George, & Prevot, (2011)		X		X					X
Mazza & Dimitrova (2007)		X							X
McNely, Gestwicki, Hill, Parli-Horne, & Johnson (2012)		X		X	X	X			
Merceron & Yacef (2005)		X			X				X
Nedungadi & Raman (2012)		X			X				X
Nelson, Bowman, & Bowman (2017)		X							X
Papamitsou & Economides (2015)		X		X					X
Scheuer & Zinn (2007)		X			X				X
Sewell, Morris, & Blevins (2008)		X			X				
Shirley & Irving (2015)		X	X				X		
Wang (2008)	X			X			X		

Figure 1. Reviewed articles categorized by data type, data processing and feedback presentation.

Data Types

While the types of data found in the review actually more closely resemble a continuum, we grouped the results into two categories for the sake of analysis.

Machine Scored Data. This data type were most readily represented by objectively scored assessment items. These items allowed assessments to be easily scored because they only allow a finite amount of possible answers to a given assessment item. An early published example of this comes from Buchanan (1998) whose PsyCal tool used objectively scored items as a formative assessment tool for students. Cassady, Budenz-Anders, Pavlechko, and Mock (2001) used Quiz Editor JS, an online assessment tool that easily creates objectively scored assessments that provide formative feedback. Additional examples of machine graded data in formative assessment can be seen in Henly (2003), Lin and Lai (2014), Wang (2008), and Alemán, Palmer-Brown, and Jayne (2011).

Activity Stream Data. As previously mentioned, the shift from machine scored to activity stream data is an artificial categorization placed on a continuum. Therefore, some studies using activity stream data are similar to those in the machine scored section. In these studies, objectively scored assessment data were collected along with student course activity data. For example, Heift (2005) developed a web-based tool for teaching German that collected objectively scored assessment data along with information regarding the context of how those answers were submitted (e.g. the student's ID, the time stamp, the task id, the system feedback, and student navigation patterns).

Similar data were collected from studies investigating the use of *intelligent tutoring systems* (ITSs). Feng, Heffernan, & Koedinger (2009) reported on the use of their ASSISTment tutoring tool, which was designed to teach and assess mathematics to secondary education students. In their system, student data included the percentage of items that were answered correctly (machine scored data), number of items completed, total time spent on the items, and the number of hint requests (in ITSs, students can typically ask for a hint when moving through a problem step). Similarly, Chen & Chen (2009) collected student interactions within a closed web-based learning environment. Logged data included correct response rate, reading rate of instructional materials, reading time, effort level of studying course materials, final test grade, attendance rate, accumulated score of question and answer and concentration degree.

This trend of collecting log data from student activities in a computer-based online environment continues in many other studies. May, George, and Prevot's (2011) Track Analysis and Visualization tool (TrAVis) logged time spent, connection frequency, message activity, and discussion threads started. Ali, Hatala, Gašević, & Jovanović's (2012) LOCO-Analyst collected student visits on certain lessons and the estimated difficulty of those lessons. In addition, the tool collected social data, including the number of sent/received messages in the course forums and chat rooms. Similar student activity streams were collected in several of the other studies reviewed (Hogaboam et al., 2016; Jin, 2017; Kosba, Dimitrova, & Boyle, 2007; Mazza & Dimitrova, 2007; Merceron & Yacef, 2005; Nedungadi & Raman, 2012; Papamitsiou & Economides, 2015; Scheuer & Zinn, 2007).

McNely, Gestwicki, Hill, Parli-Horne, & Johnson (2012) took an unconventional approach to capturing activity stream data outside of a learning environment. They created Uatu, a small program that collects collaboration information from a Google Docs document. Once Uatu is added to a created document, it logs all of the revisions that are made to that document. Their hope was that logging the revision activity would act as a metacognitive tool for its users.

While several of the previously mentioned studies collected data regarding social activity in the class, the data collected was fairly high level (frequency counts, who the message was sent to). Hsu & Ho (2012) took a further step into gathering activity stream data by collecting text content from students' discussion board posts. Their goal was to find ways of automating the feedback process. Leeman-Munk, Wiebe, & Lester (2014) also sought to automate the process of giving feedback to text inputs, but within an assessment environment. They collected short text answers from Leonardo, a virtual environment used to teach science to upper elementary students, and used Write Eval to analyze the text responses. Nelson, Bowman, & Bowman (2017) also captured text inputs from museum visitors, but in the form of museum visitors.

Real-time activity streams. The most unstructured data encountered in our review could be classified as real-time activity streams. These were moment-by-moment descriptions of student activities, often presented in real-time. The most basic example of this is seen in Shirley & Irving's (2015) investigation into the use of Connected Classroom Technology (CCTs) or audience response systems. In their qualitative investigation of the use of these tools, the

authors described four middle and high school science classrooms that use TI-Naviator™, a system that collects and displays the real-time activity stream of students. This system displays exactly what a student has typed into their own calculator, allowing the students to compare their responses with fellow students.

Other studies took on more complex data from student activity streams. Koh, Basawapatna, Nickerson, & Repenning (2014) collected the activity streams of middle school students who were engaged in the process of designing electronic games using the REACT system (Real Time Evaluation and Assessment of Cognitive Thinking). While not specific about what data were collected from the student process, the authors mention that the system "breaks down all collectable student project information and records it in the REACT database" (p. 51). Therefore, it appears that REACT is collecting a real-time activity stream of a student's game development process.

Finally, the most complex real-time activity stream data were collected by another pair of studies (Kennedy, Ioannou, Zhou, Bailey, & O'Leary, 2013; Sewell et al., 2008). Here the researchers collected activity streams from students working with a 3D dental simulation tool. In the study, subjects manipulated two pen-like haptic tools (that simulate a drill and an irrigator) in a 3D virtual space. While the students are working, the tool simultaneously recorded 48 metrics generated in real-time. Notable metrics included timestamp, tool position and orientation, the size and shape of the tool, and information about the anatomical structure they are working with. This detailed data could then be played back in real-time to give an account of the student's procedure.

Thus far we have looked at the spectrum of data being collected by these systems for the purpose of formative assessment. Data in its most structured form was machine scored assessment data. As the data becomes less structured, we saw systems that collected activity streams of students, sometimes in real-time. Next we will report how this collected data were processed for the purposes of formative assessment.

Data Processing Methods

In our next question, we wanted to understand methods used to process data to generate feedback. Here processing refers to means by which data is retrieved and analyzed from a database. Three main data processing methods categories emerged from our review: First, we had data that was unprocessed (activity stream data). Second, we had data that was produced from descriptive data analysis. Lastly, we have data that was processed through the use of data mining techniques.

Activity Streams. Data that is presented as activity streams are completely unprocessed. In our review, we found only one example of this—Shirley and Irving's (2015) research into the use of Connected Classroom Tools. In their study, the TI-Navigator™ system would directly display the problem solving processes of individual students on the main screen for everyone in the class to see. This type of data processing might be used in other classroom-response system scenarios.

Descriptive Data Analysis. The next set of results from types of data processing methods consists of descriptive student data compiled from simple database queries—in the form of counts, sums, or averages. An example of this is found in Cassady et al., (2001), where students took short formative quizzes authored in Quiz Editor JS. Data was then processed using simple descriptive statistics—counts of the number of questions that were right or wrong. Heift's (2005) German language tutor presented objectively scored assessment as well as tracking data with student users through queries. Performance data was in the form of percentages correct, as well as counts of different types of errors (spelling, verb inflection, and word order). The use of these types of simple queries to process and present data is common in many instructional systems that keep a log of student assessments and other activities (Buchanan, 1998; Feng et al., 2009; Henly, 2003; May et al., 2011; McNely et al., 2012; Papamitsiou & Economides, 2015; Wang, 2008). In order to identify commonly queried topics in a museum environment, Nelson et al. (2017) used simple counts that were then fed to a dashboard.

Data Mining. When data collected is as an activity stream, many researchers turn to data mining for processing. Data mining is used when answers to questions cannot be found simply by descriptive data analysis alone, but rather hidden patterns in the data need to be discovered (Dunham, 2003). Data mining has risen within the past few years as a topic of increasing interest to the educational community (Baker, 2010).

In this review, we found several different methods of data mining to process activity stream data collected in a technology-enhanced learning environment. While the methods are diverse, we will group them into three categories based on the goal of the method, including: Building a model of student performance, creating an expert model for automated feedback, and making recommendations for remediation.

Building a model of student performance. The goal of data processing methods in this category is to help the instructor or student understand the student's current level of knowledge in a particular domain. While a specific method wasn't stated, Koh et al. (2014) used data mining techniques to parse through the activity streams of students participating in game programming. Their goal was to understand the level of performance on a game design task.

Merceron and Yacef's (2005) TADA-Ed tool modelled student performance through several data mining techniques— k-means, hierarchic clustering, and decision trees. K-means and hierarchic clustering are both ways of grouping students or student responses together using some criteria, while decision trees are a way to classify an object based on other similar objects. Hogaboam et al. (2016) used similar techniques in their investigation of a tool designed for problem-based learning. Using a mixture of descriptive analysis and data mining, they were able to identify areas of student misunderstanding.

In another tool for teachers, Kosba et al. (2007) used a fuzzy student modeling approach to generate a student's certainty factor for every concept that they covered in a learning environment. Based on these models, a concept is assigned as completely learned, learned, or not learned for a given student or student group. In addition, the system builds a model of groups within the class based on inputted criteria, e.g. nationality, background, or course preferences. These features allow instructors to track the progress of individual students in the class as well as selected groups.

Alemán et al. (2011) processed objectively scored data using Snap-Drift Neural Networks, a high-speed data categorization algorithm, to group similar responses together. The resulting groups consist of students making the same mistakes on the same problems. The authors felt that these groups represented students who had similar misconceptions that an instructor might be able to correct through remediation. Similarly, Lin and Lai (2014) used an a priori algorithm to identify questions on an objectively scored assessment that were commonly missed together.

To model student knowledge to the students themselves, J. L. Hsu, Chou, and Chang (2011) used text mining methods to provide students with feedback regarding the quality of their answers in an online discussion board. Using Latent Semantic Analysis and multiclass singular value decomposition, their tool automatically classified student responses to the discussion board according to its depth of understanding as described by Bloom's taxonomy. Jin (2017) also used discussion activity to perform social network analysis, to help students become more aware of their performance within a group.

Creating an expert model for automated feedback. The next set of studies used data mining to build models of expert performance, which could then be used to generate automatic feedback to students. To create automated feedback to constructed text response data in an online science learning environment, Leeman-Munk et al. (2014) used machine learning techniques. Machine learning is concerned with "design[ing] systems that can learn from data" (Bell, 2014, p. 2). They trained their system on sample human-graded correct and partially correct answers and then used soft cardinality and Latent Semantic Analysis to automatically grade submitted answers that were not completely identical to the sample answers.

Kennedy et al. (2013) used several algorithms to give students expert feedback in a 3D dental simulator. First, they identified an association rule that could be calculated in real-time that would demonstrate surgical expertise. Next, they took this association rule and used it to build a Hidden Markov model (taking an observable metric and then inferring a hidden gesture not easily observable through those metrics alone). In this case, the Hidden Markov model would infer whether a participant was engaging in a stabbing motion, a sweeping motion, or a stabbing-sweeping motion with their dental instrument. The researchers then trained these models on the performance of expert and novice surgeons to determine when the tool should give feedback. Many of the same data analysis techniques were used by Sewell et al., (2008).

Making recommendations for remediation. In the last set of data mining methods, the goal was to provide recommendations for the future learning for students. Scheuer and Zinn's (2007) Student Inspector used machine learning "to predict future learning outcomes" (p. 6). Instructors could use *the Analyser* (a module within Student

Inspector) to suggest appropriate lessons to students based on their performance history. Poorly performing students could have easier tasks recommended to them, while higher performing students were recommended more challenging tasks.

Chen and Chen (2009) used several sequential data mining techniques to make learning recommendations in an online learning environment. First, they performed factor dependence analysis using a fuzzy clustering method to determine student measures that were most essential to the analysis. The final measures included reading rate, correct response rate of test items, accumulated score on the discussion board, and effort level of studying course materials. Second, fuzzy association rule mining was used to create rules that would predict student performance. One example of a rule is "CR_L => GRADE_L", interpreted as *if a student has a low correct response rate on a test, they are likely to have a low cumulative grade*. The system could then use these rules to provide formative feedback to students in the course.

The studies above demonstrate that there are a variety of data analysis techniques currently being employed to process both machine gradable and activity stream data from online learning environments. In the next section, we will discuss how this information, once retrieved, is being presented to the instructor or student for the purposes of formative assessment.

Data Presentation for Formative Assessment

Once processed, data from learning systems can be presented to both students and instructors for the purpose of providing feedback on the learning process. In this section, we will first talk about how data were presented to students for the purpose of formative assessment, and then we will discuss how data were presented to instructors.

Students. After reviewing the literature, we discovered two main methods of delivering feedback to students based on data collected in the learning process. The first is through the use of student reports and dashboards and the second is through real-time feedback delivered to the student while engaged in the learning activity.

Student Reports and Dashboards. The first category of feedback delivery to students is in the form of reports and dashboards. These are both visual displays of student activities that can be viewed directly by the student—the report is in a tabular format, while the dashboard includes graphic representations of the data. In our review, the first report was shown in Buchanan (1998), whose PsyCal systems was designed to be a formative assessment tool for an Introduction to Psychology course. Once a student submitted their answers to an assessment, they were presented with a report that told them how many questions they answered correctly, along with a list of questions that were incorrectly answered. Correct answers were not given, but rather the student was referred to sections of the text where correct answers could be found.

After students completed learning activities in Heift's (2005) German language tutoring system, they viewed their progress via The Report Manager which kept detailed error reports, identifying which items on activities or assessments were missed. For those exercises in which the student wished to achieve a better score, the Report Manager allowed them to redo exercises. In an evaluation of the effectiveness of the Report Manager, Heift (2005) found that of users who viewed a learning report, 70% of them repeated whole exercise sets after viewing the results – suggesting that viewing student performance levels could be a trigger for seeking better performance in a class. A similarly styled learning report was also used by Nedungadi and Raman (2012) to give feedback to students after engaging in a mobile learning session, and Lin & Lai (2014) after answering formative questions in their annotation-sharing and intelligent formative assessment (ASIFA) system. In examining this set of reports, it would seem that the most useful are those that encourage some kind of action on the part of the student. This may include revisited material not understood, or working on a problem set that wasn't previously mastered.

McNely et al. (2012) were particularly interested in the metacognitive value of a report for students when they developed Uatu—a tool which tracks contributions to a Google document. Their hope was that by seeing a visual display of how group members had contributed to the document, they would be more likely to increase their own participation. However, the authors did not perform any evaluation on the tool, and so it is difficult to know whether the tool had the intended impact.

Jin (2017) also used student facing visualization for metacognitive purpose. In the study, students were shown a social network analysis graph of their participation in a course. The author tested two different ways of displaying student and group activity. According to the results of their experimental design, both methods increased the perception of student performance over control conditions.

Hsu et al. (2011) wanted to provide students with a more visual display of their performance of a learning environment. After using text-mining algorithms to classify student discussion responses by their level of understanding on Bloom's cognitive taxonomy, the system produced two graphs—one for individual cognition level, and one for the cognition level of the class. The individual cognition displayed a spider (or radar) chart with the student's scores for each of the six cognition levels on a scale from 0 to 1. If a student had a lower score for a certain cognitive level, then they had failed to use words that correctly corresponded with that level. In the class level graph, the student could see common words used by other students in their class and how those words were assessed in terms of cognitive understanding. This graph gave feedback on the types of words that the student should use to increase their cognitive understanding of the underlying concept. While Hsu et al.'s (2011) system is conceptually interesting, it doesn't seem to directly address some large assumptions, namely the validity of using words and phrases to assess underlying cognitive understanding. We feel that using phrases that match a pre-determined level of understanding is not valid evidence that a student is actually thinking at that level.

In a final study where feedback was directed specifically to students, Chen and Chen (2009) used two reports to help learners see their progress in an online learning module. In the first report, students were presented with several metrics gathered from the learning system and the instructor. These included attendance rate, concentration rate, instructor comments, and correct response rate. In addition, the student's final score was predicted based on the associated fuzzy rules created in the earlier data mining process. In the second report, the student was able to see the previously generated fuzzy learning rules for the class. An example of this is *High RT (Reading Time) => High Score*—meaning high reading time leads to a higher score. While such rules might be interesting for an instructor to look at, they seem inappropriate a group of 9 to 11 year old students (the sample in the study) to understand. In addition, the authors don't address whether there might be possible detrimental effects to student motivation after they are shown a predicted score.

Real-time feedback to students. In the next group of studies, feedback was delivered to students in real-time while engaged in the learning activity. This approach is seen in its most basic form by Cassady et al. (2001) in their work with embedded formative quizzes. While the authors didn't delve deeply into the type of feedback given to students, they did mention that the feedback was immediate. Because the data were objectively scored, it is likely that the feedback was pre-programmed by the assessment designer. While these systems can give feedback quickly, the type of feedback is not customized. The system knows whether or not a student got a question wrong, but it does not know the reason. The use of quick feedback following objectively scored assessment was also utilized by Henly (2003) and Wang (2008).

Another relatively simple mechanism for real-time student feedback was examined by Shirley and Irving (2015). The TI-Navigator™ tool in their study provided a real-time activity stream of student responses viewable to members of the class. Notwithstanding, in this case, the students had to derive feedback themselves from the activity stream by comparing their answers to other students in the class as the system did not process the student data.

Kennedy et al., (2013) demonstrated the most sophisticated real-time student feedback system. In their research with a 3D dental simulator, feedback was presented to the student in the form of real-time suggestions for better performance in a text box on their display. In the example provided in their article, the system detected when a student was not performing the procedure in the optimal manner. At this point a small text box appeared on the student view saying, "You are too tentative at this stage of the procedure. Apply more force" (p. 179). While their research is promising, the authors note that their real-time feedback process is in its initial stages. It is clear that more research needs to be done to understand how often such feedback could be presented before the student before they feel like they are being pestered.

Instructors. Our review essentially found one way of presenting feedback to an instructor: the report. However, these reports vary in their level of detail and their reliance on visual representations. In this section, we will briefly discuss more traditional reports and then discuss the use of visual reports (commonly referred to as dashboards).

Instructor reports. Instructor reports are tables that present data to an instructor about student performance in an online environment. In our review, reports appeared in their most simple form in a study by Feng et al. (2009), whose ASSISTment system allows instructor to run several reports, including a grade book. The grade book delivers several data points from the ASSISTment system back to the instructor in tabular form. The instructor could then look at possible patterns and intervene if necessary. Chen and Chen (2009) processed student and instructor data were delivered via a report to the instructor's mobile device. This report included learner test scores, the variances in learner scores, and learning rules derived from the earlier data mining process. While the information that is reported from this process may be useful, it would seem difficult for a typical instructor to derive meaning and take an action from such a table of numbers. This was also a problem in Aleman, Palmer-Brown, and Jayne's (2011) research which processed a series of formative, multiple-choice questions using data mining. The result of this data processing was then presented to the instructor as groups of responses to a series of questions. The authors explained:

> For example, "b/d c *" represents a group characterized by all the students answering b or d to question 1, c to question 2, and mixed answers to question 3. Hence the educator can easily see the common mistakes in the groups of the student answers highlighted by the tool (p. 503).

Once again, while the information derived from the data processing may have been useful, it still required the instructor to sufficiently understand the content of each of the questions to derive meaning from the data.

The Teach Advisor (TAdv) tool described by Kosba et al. (2007) attempted to simplify the process of turning instructor reports into actionable feedback. In the *Generate advice* and *View advice* sections of the TAdv tool, the system created advice for the instructor based on the algorithm used in the earlier data processing step. This feedback could be addressed to individual students, groups of students within the class, or the class as a whole. The instructor could review the advice, and use the report to send automatically created feedback to students. Such a move seems like step in the right direction, as the system is helping direct the instructor's attention to potential learning problems in the course.

Visual dashboards. While the instructor reports in the previous section typically relied on data delivered to the instructor via a table, reports in the instructor dashboard category rely on visual representations of data. Bajzek et al. (2007) provide a good definition of a dashboard when talking about their own Digital Dashboard for Learning (DDL) for Carnegie Mellon University's Open Learning Initiative (OLI). According to them, a dashboard "provides visibility into key indicators of student learning through simple graphics such as gauges, charts, and tables within a web browser" (p. 2). The authors then gave several key components to an effective educational dashboard:

- Provide a wide variety of different metrics in a single consolidated view
- Roll-up details into higher level summaries
- Provide intuitive visualizations that are instantly understandable – for example, red bars mean a problem
- Provide linkages to the data that they represent. (p. 3)

In one of the earliest studies using visual displays for the purpose of formative assessment, Merceron and Yacef (2005) built a tool that allowed for the cleaning, processing and visualization of educational data (TADA-Ed). While several visualizations were available for the instructor to look at, the visualizations could only be accessed once an advanced statistical analysis was run (such as a cluster analysis). While there is definitely power in such an approach, it is difficult to determine whether such complex visualizations would be helpful to the typical instructor monitoring of the learning of their class.

Mazza and Dimitrova (2007) used several visual representations in their CourseVis tool to help instructors understand how university students were progressing in an online course. In one of the visualizations, discussion forum activity was visualized through the use of a *discussion plot*. In this plot, the discussion posts originator, time of posting, and post follow-up activities are plotted on a two dimensional axis, with point size representing the number of follow-up posts. This *discussion plot* can help the instructor identify students that are making active contributions to the discussion forums. Another visualization discussed is what the authors refer to as a *cognitive matrix*, in this visualization student performance for each of the course topics in the class was visualized in a grid, with colors representing performance. This matrix enables the instructor to quickly ascertain the class' knowledge on a given topic, or a single individual's performance across a range of topics. The last described visualization is the *student*

access plots that visualize student activity in the course over time. In this set of graphics, overall logins to the course along with activity in certain content pages were graphed. The use of these *student access* plots may help an instructor gauge a student's engagement with the course.

In Scheuer and Zinn's (2007) Student Inspector, course activity data were visualized through a combination of windows that allow instructors to sort, filter, and visualize data. In the *performance measurement* tab of the program, the instructor could identify the performance of individual students compared with their peers. In the *misconceptions* tab, instructors were able to look at those topics for which students struggled the most with, the distribution of which was visualized as a pie chart. The *misconceptions* tab allowed the instructor to quickly see where they may need to remediate for individual students, or adjust instruction for the entire class. In the last visualization, the *topic coverage* tab, the instructor was able to see weak and strong performance areas for each student. Such a view also may be helpful to an instructor in identifying areas for student remediation. Overall, the Student Inspector tool seems to be useful as a tool for data enabled formative assessment. It focuses on providing the instructor with actionable information in the form of areas for student remediation. In fact, after Scheuer and Zinn's (2007) description of their product, they also provide some evaluation data. Overall, instructor reaction was positive and they felt that the most useful aspect of the program was the identification of student misconceptions.

A windowed interface similar to Student Inspector was used by Ali et al. (2012) in their LOCO-Analyst tool. In the latest version of the tool, student information is presented to the instructor in four tabs: forums, chats, learning, and annotations. The forums and chats tabs provided visualization to the instructor regarding the students' online interactions. The number of postings for each topic are presented via tabular and bar chart format. In the learning tab, the instructor is presented with bar charts regarding student activity time on topics in the course. The researchers also performed an evaluation on the tool to get feedback from potential users. Overall, response to the tool was positive, however, most of the instructors (70%) felt that the tool did "not provide enough information on how to improve student's online interactions" (p. 482). This suggests that there is something about the type of data being presented to the instructors that is not actionable.

May et al.'s (2011) Tracking Data Analysis and Visualization (TrAVis) tool used two visualization tools to give information to instructors. The first, the *Time Machine,* allows instructors to see a timeline view of completed student activities. The second tool, a radar graph, "provid[es] simultaneous observation and analysis of different aspects of user activity" (p. 61). In one example, the radar graph visualizes connection frequency, discussion threads started, messages posted, messages replied, and messages quoted. These radar tools are also applied to group discussion, where an instructor can quickly see online group activity at a glance. While these visualization tools may be good for instructor awareness of course activity, they also offer little by way of actionable information to an instructor. A similar approach to instructor visualization was utilized in a study by Hogaboam et al., (2016) and Papamitsiou & Economides (2015). Nelson, Bowman, & Bowman, (2017) differed slightly in that the end audience were museum staff, but the visualizations were similar to those used by an instructor.

A radar graph was the primary visualization in Koh et al.'s (2014) Real Time Evaluation and Assessment of Computational Thinking (REACT) tool. In their Computational Thinking Pattern Analysis Graph, student activities were mapped onto nine different *computational thinking patterns* (cursor control, generations, absorptions, collision, transportation, push, pull, diffusion, and hill climbing). Individual student graphs are then aggregated on the Assessment Dashboard and given color codes to indicate how well a student is progressing through their programming process. Green graphs indicate that the student is working correctly, while orange and red graphs suggest to the instructor that a student may be struggling. In testing the tool with a sample of instructors, the response was positive, and the instructors felt that the visualizations helped them keep tabs on the work that their students were doing.

Discussion

The aim of this research was to review current practices in data-enabled formative assessment. In doing so, we wanted to understand the types of data being collected, the way that this data were being processed, and how the processed data were presented before students or instructors for the purpose of formative assessment.

While a few systems were collecting only objectively scored data (i.e. multiple choice and true-false questions), a majority of the systems were collecting additional data about the users of these systems. While objectively scored data were convenient for giving pre-programmed feedback to the student (as seen in Cassady et al., 2001), it cannot give

targeted feedback to the student or make inferences about why the student responded incorrectly to the question. To give more nuanced feedback to students, additional information needs to be collected. This was demonstrated by the many studies that collected data from student activities. By collecting information about assessments, social activity, and course usage, the systems had more information to make recommendations or suggestions for remediation to the instructor or the student. Additionally, by keeping a log of a student's progress throughout a course, the systems were able to see larger trends for an individual student, and identify to instructors that students may be demonstrating that they are struggling with material.

When it comes to processing data, many of the systems that we reviewed used descriptive data analysis to collect information and present it to the instructor. These systems query student logs in the database to create reports for instructors and students. While such queries are useful, they may not be uncovering information that would normally be hidden to an instructor or student. To uncover these more nuanced patterns, other systems used data mining techniques. Overall, there was not one clear data mining method being used, it largely depended upon the nature of the data and the end goal. A few of the products that we reviewed (especially Merceron & Yacef, 2005) relied on the instructor to pick the appropriate data mining method. While data mining has proven to be an effective way of finding patterns in student data, we cannot expect instructors to be experts in its method. Because of this, most of the actual processes should be hidden to the instructor (with possible advanced option for instructors who want additional control). As data mining processes become more common in educational research, more effort needs to be given to explain these methods to non-experts so that appropriate research questions can be formulated.

Once this educational data were processed, we saw several ways that it was presented to both students and instructors. For the students, several studies found value in using reports and dashboards for meta-cognitive purposes (Heift, 2005; McNely et al., 2012; Nedungadi & Raman, 2012). However, these approaches gave students high-level feedback and not specific areas for remediation. A review of Chen and Chen (2009) also pointed to the need to make recommendations appropriate to the age level and ability of the students—presenting the model directly to the student does not seem like the best approach. Also unknown is the motivational effect of presenting such models directly to students. More research needs to be focused on understanding how best to give feedback to students in these environments. It will also be important for research to look more into appropriateness and practicality of real-time feedback to students as is demonstrated by Kennedy et al., (2013). Such feedback may be more appropriate for simulations (as in surgical training), but it would be interesting to see it applied to other contexts. For many of these issues, we may want to look to intelligent tutoring literature to see what it has uncovered about feedback in learning environments.

We also discovered many ways that data from these learning environments is being presented back to instructors for the purpose of formative assessment. The first category in these types of feedback was instructor reports. While such reports are likely useful to instructors, they may make it difficult for instructors for identify important patterns in student learning that need to be corrected. Instructor dashboards were more helpful in accomplishing this goal. Much like data mining techniques, the designs of these dashboards varied from project to project, depending upon the type of data that is being collected and the specific needs of the instructor. We should place more effort on developing customizable instructor dashboards that present a variety of choices for displaying student data.

It should be noted that the most salient feature of the dashboard identified in Scheuer and Zinn (2007) was the ability to identify student misconceptions. This points to the question of actionability of the data in these visual dashboards for instructors. More research needs to be focused on identifying the utility of dashboard features for instructors. What data are important for an instructor in a dashboard? What presented information are they already aware of? What is going to help them make pedagogical decisions?

Implications for Future Research

Through this review, we've identified several areas where further research could help us develop smarter tools to enable formative assessment. First, as more and more educators are looking to use data processing and mining techniques to find patterns in student data, we need more accessible materials to help them understand these processes. We hope that in the future, researchers will provide better rationales for their use of data analysis techniques depending on the type of educational data at hand. Currently, much of the research in the fields of Learning Analytics and Educational Data Mining is being created by computer scientists. Making these methods more accessible to education researchers and practitioners could allow them to contribute more to the direction of the field.

Our review also suggests that we need to look more at how students and instructors are using reports and dashboards. Our critique of Chen and Chen's (2009) research points out the sometimes inappropriate nature of dashboard feedback to students. Other tools that we have identified in our study (Merceron & Yacef, 2005) seemed too complex for a normal instructor to interpret. More empirical research and evaluation on the use of these tools could help designers create visualizations that are appropriate to the development level of the user and also helpful in producing specific recommendations.

A final area of research arises from a finding of Scheuer and Zinn (2007). In the evaluation of their tool, the ability to identify learner misconceptions was seen as the most helpful feature to instructors. This finding doesn't surprise us, as we feel that instructors want to use dashboards as tools to understand areas of their class that they can improve—the identification of misconceptions gives a concrete starting point for a remediation in their class. We feel that more research is needed regarding how we can use these data-enabled tools to identify misconceptions among students, especially misconceptions that might not be present in larger grained data.

Conclusion

The goal of this review was to review tools that have been used to facilitate formative assessment through the use of computer collected data. From the amount of research that we have found, it appears that many in the fields are making great strides in harnessing the power of these tools to help give instructors information about student performance. The most effective of these systems collect as much student performance data as possible, parse through the data using advanced analysis techniques, and then present patterns and trends back to the instructor or teacher using visual techniques.

We hope that this review will act as a starting point for future designs of online learning and assessment tools. Data can be a powerful tool, which when applied appropriately can give instructors great insight into the learning process. These insights will hopefully allow them to do what they are best at—teaching.

References

Alemán, J. L. F., Palmer-Brown, D., & Jayne, C. (2011). Effects of response-driven feedback in computer science learning. *IEEE Transactions on Education*, *54*(3), 501–508. https://doi.org/10.1109/TE.2010.2087761

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, *58*(1), 470–489. https://doi.org/10.1016/j.compedu.2011.08.030

Bajzek, D., Brown, W., Lovett, M., & Rule, G. (2007). Inventing the digital dashboard for learning. *World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, 1084–1092. Retrieved from http://www.editlib.org/p/25512

Baker, R. S. J. D. (2010). Data mining for education. In *International Encyclopedia of Education* (Vol. 7, pp. 112–118). https://doi.org/10.4018/978-1-59140-557-3

Bell, J. (2014). *Machine Learning: Hands-On for Developers and Technical Professionals*. Indianapolis, IN: John Wiley & Sons.

Buchanan, T. (1998). Using the World Wide Web for Formative Assessment. *Journal of Educational Technology Systems*, *27*(1), 71–79. https://doi.org/10.1016/s0360-1315(00)00049-x

Cassady, J. C., Budenz-Anders, J., Pavlechko, G., & Mock, W. (2001). The effects of internet-based formative and summative assessment on test anxiety, perceptions of threat, and achievement. In *Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001)* (p. 13).

Chen, C.-M., & Chen, M.-C. (2009). Mobile formative assessment tool based on data mining techniques for supporting web-based learning. *Computers & Education*, *52*(1), 256–273. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2008-17099-026&site=ehost-live&scope=site

Dunham, M. H. (2003). *Data Mining: Introductory and Advanced Topics*. Upper Saddle River, New Jersey: Prentice Hall.

Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, *19*(3), 243–266. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2009-10369-003&site=ehost-live&scope=site

Hattie, J. (1999). Influences on student learning. *Inaugural Lecture given on August*, *2*, 1–25. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.114.8465&rep=rep1&type=pdf

Heift, T. (2005). Inspectable learner reports for web-based language learning. *ReCALL*, *17*(May 2005), 32–46. https://doi.org/10.1017/S0958344005000418

Henly, D. C. (2003). Use of Web-based formative assessment to support student learning in a metabolism/nutrition unit. *European Journal of Dental Education : Official Journal of the Association for Dental Education in Europe*, *7*(3), 116–122. https://doi.org/10.1034/j.1600-0579.2003.00310.x

Hogaboam, P. T., Chen, Y., Hmelo-Silver, C. E., Lajoie, S. P., Bodnar, S., Kazemitabar, M., & Wiseman, J. (2016). Data dashboards to support facilitating online problem-based learning. *The Quarterly Review of Distance Education*, *17*(3), 75–91.

Hsu, C.-C., & Ho, C.-C. (2012). The design and implementation of a competency-based intelligent mobile learning system. *Expert Systems with Applications*, *39*(9), 8030–8043. https://doi.org/10.1016/j.eswa.2012.01.130

Hsu, J. L., Chou, H. W., & Chang, H. H. (2011). EduMiner: Using text mining for automatic formative assessment. *Expert Systems with Applications*, *38*(4), 3431–3439. https://doi.org/10.1016/j.eswa.2010.08.129

Jin, S.-H. (2017). Using Visualization to Motivate Student Participation in Collaborative Online Learning Environments. *Educational Technology & Society*, *20*(2), 51–62.

Kennedy, G., Ioannou, I., Zhou, Y., Bailey, J., & O'Leary, S. (2013). Mining interactions in immersive learning environments for real-time student feedback. *Australasian Journal of Educational Technology*, *29*(2), 172–183.

Koh, K. H., Basawapatna, A., Nickerson, H., & Repenning, A. (2014). Real time assessment of computational thinking. In *2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 49–52). IEEE. Retrieved from http://sgd.cs.colorado.edu/wiki/images/9/91/Paper_24.pdf

Kosba, E., Dimitrova, V., & Boyle, R. (2007). Adaptive feedback generation to support teachers in web-based distance education. *User Modelling and User-Adapted Interaction*, *17*, 379–413. https://doi.org/10.1007/s11257-007-9031-z

Leeman-Munk, S. P., Wiebe, E. N., & Lester, J. C. (2014). Assessing elementary students' science competency with text analytics. In *Proceeding of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 143–147). ACM. https://doi.org/10.1145/2567574.2567620

Lin, J. W., & Lai, Y. C. (2014). Using collaborative annotating and data mining on formative assessments to enhance learning efficiency. *Computer Applications in Engineering Education*, *22*(2), 364–374. https://doi.org/10.1002/cae.20561

May, M., George, S., & Prévôt, P. (2011). TrAVis to enhance online tutoring and learning activities: Real-time visualization of students tracking data. *Interactive Technology and Smart Education*, *8*, 52–69. https://doi.org/10.1108/17415651111125513

Mazza, R., & Dimitrova, V. (2007). CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human Computer Studies*, *65*, 125–139. https://doi.org/10.1016/j.ijhcs.2006.08.008

McNely, B. J., Gestwicki, P., Hill, J. H., Parli-Horne, P., & Johnson, E. (2012). Learning analytics for collaborative writing: a prototype and case study. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 222–225. https://doi.org/10.1145/2330601.2330654

Merceron, A., & Yacef, K. (2005). TADA--Ed for Educational Data Mining. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, *7*(1), 267–287. Retrieved from http://imej.wfu.edu/articles/2005/1/03/printver.asp

Nedungadi, P., & Raman, R. (2012). A new approach to personalization: Integrating e-learning and m-learning. *Educational Technology Research and Development*, *60*, 659–678. https://doi.org/10.1007/s11423-012-9250-9

Nelson, B. C., Bowman, C., & Bowman, J. (2017). Designing for Data with Ask Dr. Discovery: Design Approaches for Facilitating Museum Evaluation with Real-Time Data Mining. *Technology, Knowledge and Learning*, *22*(3), 427–442. https://doi.org/10.1007/s10758-017-9313-4

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology and Society*, *17*(4), 49–64.

Papamitsiou, Z., & Economides, A. A. (2015). Temporal learning analytics visualizations for increasing awareness during assessment. *RUSC. Universities and Knowledge Society Journal*, *12*(3), 129. https://doi.org/10.7238/rusc.v12i3.2519

Sadler, D. R. (1998). Formative Assessment: revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 77–84. https://doi.org/10.1080/0969595980050104

Scheuer, O., & Zinn, C. (2007). How did the e-learning session go? The Student Inspector. *Proc. of the Conf. on Artificial Intelligence in Education (AIED'07)*, 487–494. Retrieved from http://dl.acm.org/citation.cfm?id=1563601.1563678

Sewell, C., Morris, D., Blevins, N. H., Dutta, S., Agrawal, S., Barbagli, F., & Salisbury, K. (2008). Providing metrics and performance feedback in a surgical simulator. *Computer Aided Surgery*, *13*(2), 63–81. https://doi.org/10.1207/s15327752jpa8502

Shirley, M. L., & Irving, K. E. (2015). Connected classroom technology facilitates multiple components of formative assessment practice. *Journal of Science Education and Technology*, (24), 56–68. https://doi.org/10.1007/s10956-014-9520-x

US Department of Education. (2011). *Transforming American Education: Learning powered by technology* (Vol. 8). https://doi.org/10.2304/elea.2011.8.2.102

Wang, T.-H. (2008). Web-based quiz-game-like formative assessment: Development and evaluation. *Computers & Education*, *51*(3), 1247–1263. https://doi.org/10.1016/j.compedu.2007.11.011