JRTE | Vol. 47, No. 2, pp. 89–104 | © 2015 ISTE | iste.org/jrte
DOI: 10.1080/15391523.2015.999639

Toward a New Approach to the Evaluation of a Digital Curriculum Using Learning Analytics

Virginia Snodgrass Rangel
Third Coast Insight

Elizabeth R. Bell
Accelerate Learning Inc.

Carlos Monroy
Rice University

J. Reid Whitaker
Accelerate Learning Inc.

Abstract

Understanding how an educational intervention is implemented is essential to evaluating its effectiveness. With the increased use of digital tools in classrooms, however, traditional methods of measuring implementation fall short. Fortunately, there is a way to learn about the interactions that users have with digital tools that are embedded into the technologies themselves: user data. The purpose of this article is to outline ways in which researchers can harness learning analytics and user data to gain a deeper understanding of the implementation of digital innovations and their impact on teaching and learning. We discuss four considerations for the integration of learning analytics and user data in implementation research, and we provide an example of integration from an evaluation of a digital science curriculum, STEMscopes. (Keywords: implementation research, learning analytics, digital learning)

Understanding how an educational intervention is implemented in the classroom is essential to evaluating its effectiveness (Century, Cassata, Rudnick, & Freeman, 2012; Durlak & DuPre, 2008; Franks & Schroeder, 2013; Hagermoser Sanetti & Kratochwill, 2009; Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012; Song & Herman, 2010). Recent interest in implementation is due in large part to state and federal requirements that schools implement evidence-based educational resources (e.g., Reading First, Race to the Top, Investing in Innovation). In response to the demand for a deeper understanding of how implementation influences the effectiveness of interventions, researchers have designed new protocols for measuring the different dimensions of implementation, including surveys, logs, and observations.

With the increased use of digital and online tools in classrooms, however, traditional methods of measuring implementation fall short. Digital interventions can include computer programs, software, and websites that provide content, tutoring, assessments, simulations, and games, and often are utilized by hundreds of thousands to millions of students. For example, programs such as DreamBox Learning, A2i, Assistments, READ180, ST Math, Reasoning Mind, and Khan Academy, among many others, offer a wide variety of online activities, assessments, and even "adaptive"

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/UJRT.

algorithms, and are becoming more common in classrooms across the country. When a user (e.g., student or teacher) interacts with these programs, most to all of the interaction is between the user and the program. As such, understanding how the system is being implemented, and therefore its impact on learning, requires data about those interactions, which cannot be captured reliably via observations or surveys.

Scholarly research has struggled to keep up with the rapid proliferation of digital tools (Bienkowski, Feng, & Means, 2012; Means, Anderson, & Thomas, 2012). Much of the research that does exist documents programs developed before the most recent increase in interest and use of digital tools (Chumley-Jones, Dobbie, & Alford, 2002; Means, Toyama, Murphy, Bakia, & Jones, 2010) and does not examine differences in levels of utilization (we discuss recent exceptions later). Fortunately, there is a way to learn about those interactions that is embedded in the digital technologies themselves: user data. The analysis of educational user data has become commonplace in the rapidly growing field of learning analytics (LA), which uses diverse data sources—including learners' and educators' interaction data, data mining techniques, and theories from information sciences, psychology, cognitive science and other disciplines to understand and improve "learning and the environments in which it occurs," (Siemens, 2012, p. 1). Together with more traditional measures of implementation, user data can help create a richer picture of the implementation and impact of digital innovations.

The primary purpose of this article is to outline ways in which researchers can harness the potential of LA techniques and user data to gain a deeper understanding of how digital innovations are being implemented in and out of classrooms and their impact on teaching and learning. We also provide an example from a recent evaluation of a digital science curriculum, STEMscopes, of how user data and LA techniques can be integrated into program evaluation in education. To conclude, we discuss challenges to the integration of LA techniques and user data, as well as potential solutions.

Literature Review

Measuring Fidelity of Implementation

Best practices in measuring implementation are still debated. Generally, implementation refers to how an innovation is used (Song & Herman, 2010), and fidelity of implementation is "the extent to which an enacted program is consistent with the intended program model" (Century, Rudnick, & Freeman, 2010, p. 2).

Researchers and policymakers understand the importance of measuring implementation because many innovations do not have the intended impact, often due to the innovation not being implemented as intended—an occurrence that Franks and Schroeder (2013) refer to as "Type III error." Indeed, early research on implementation came out of evaluations of the impact of federal programs implemented in the 1960s and 1970s, as well as evaluations of large-scale curriculum reform. As part of these evaluations, researchers found that a key reason that the programs did not have their intended impact was that "local actors" (e.g., teachers) took the interventions and adapted them to their local contexts and needs instead of using them exactly as the developers intended (Berman & McLaughlin, 1976; Fullan & Pomfret, 1977; McLaughlin, 1976).

Implementation research focuses on what happens when an innovation enters a community, school, or other organization. There are two components to the study of implementation: fidelity of implementation and the process of implementation (Fullan & Pomfret, 1977; Scheirer & Rezmovic, 1983). Fidelity addresses the organizational and curricular changes that do or do not take place as intended by the new innovation, while the implementation process refers to the ways in which local agents adapt the innovation.

The focus of this article is on fidelity of implementation (FOI). As recent research has noted, there is a growing amount of literature on "implementation science," or the measurement of FOI, with many authors giving different names to very similar constructs (Century et al., 2012; Nelson et al., 2012). Here, we briefly highlight similarities across some of this work to focus on where the use of LA techniques and user data may be appropriate.

Table 1. Traditional Measures of Fidelity of Implementation

Adherence	Dosage	Quality	Participant Engagement
Teacher Self-Reports Observations	Teacher Logs or Checklists Extrapolations based on observations of a sample of lessons Attendance data	Teacher Self-Reports Observations	Surveys Ways to rate "the extent to which participants are engaged by and involved in the activities and content of the program" (244)

Researchers agree that FOI is a multidimensional construct. Dane and Schneider (1998) were among the first to outline what these dimensions were and how to measure them. They defined FOI generally as "program integrity" and argued that it had the following five dimensions: adherence, exposure/dosage, quality of delivery, responsiveness of participants, and program differentiation. Since then, other researchers have built on this foundation, with some changes being small and others more substantive. In our review, we found a great deal of overlap across much of it, and in attempting to delineate where LA techniques and user data may fit, we try to draw parallels across different work and terminology.

There are several ways to gather data to measure the multiple dimensions of FOI. Dane and Schneider (1998), for example, proposed the methods summarized in Table 1 as potential tools for data collection. Similarly, if one reviews recent program evaluations, several different kinds of data collection methods make an appearance: teacher checklists, observation protocols, teacher surveys, and even student interviews or surveys (Quint et al., 2013).

Traditional methods of data collection are necessary but not sufficient for understanding the implementation of digital classroom tools. Traditional methods cannot capture a complete picture of how a digital device is being used because much of the use occurs between a user and the system/device, and therefore is not easily observable, if observable at all. Research is beginning to incorporate new measures of use to account for this change. For example, in an evaluation of the Acuity Data System, Wayman and colleagues (2009a, 2011) integrated teacher use logs as a measure of implementation. Similarly, in an evaluation of the online math program DreamBox Learning, researchers examined the number of hours students were actively working on the program, data they were able to pull from the program itself (Wang & Woodworth, 2011). In a study of the implementation of an online science inquiry program, researchers utilized student participation data that the program collected as part of its design (Penuel & Means, 2004). Other studies also have considered the relationship between usage and outcomes (Doan, Zhang, Tjhi, & Lee, 2011; Loftus, 2012; Turner, 2010; Yu & Jo, 2014). However, none of these examples used digitally generated data in a rigorous way by articulating how these data fit within current frameworks of FOI and none utilized LA techniques to uncover patterns of use.

The purpose of this article is to sketch out a more disciplined approach to incorporating LA techniques and user data into the study of FOI and evaluation of digital educational innovations. First, we do this by outlining considerations for working with LA techniques and user data. Then we build on our past work experimenting with different LA techniques (Monroy, Rangel, & Whitaker, 2013, 2014) and develop new variables that are meant to measure different kinds of utilization.

Learning analytics research. Learning analytics is an emerging field of research that applies statistical methods, theories, and models to the analysis of large and diverse educational data sets (Bienkowski et al., 2012). In education specifically, researchers can analyze different kinds of user data to make inferences about how teachers and students are interacting with technology. Here we briefly describe two current areas of research within the realm of LA and educational data mining.

The first area, which is of great interest to educators, is the impact of an intervention on student outcomes. This kind of research is more common in higher education to help retain students and improve outcomes (Dietz-Uhler & Hurn, 2013; Olmos & Corrin, 2012; Smith, Lange, & Huston, 2012). In grades K–12, researchers use student data to predict future achievement, as is the case with the ASSISTments initiative at Worcester Polytechnic Institute, which has produced a corpus of LA research ranging from measuring ASSISTment's effect on math achievement (Koedinger, McLaughlin, & Heffernan, 2010) to comparing the effect of traditional and computer-supported homework on student outcomes (Mendicino, Razzaq, & Heffernan, 2009). Similarly, using data from users interacting with videos in the Khan Academy, Muñoz-Merino, Ruipérez, and Delgado (2013) proposed a set of variables to measure diverse learning constructs. Data on student interactions also can uncover misconceptions about content and can be used to create sophisticated algorithms that recommend appropriate activities and interventions to maximize achievement (Corbi & Burgos, 2014; Khribi, Jemni, & Nasraoui, 2015; Özyurt, Özyurt, Baki & Güven, 2013).

A second area of research has used LA to gain insight into student thinking and behaviors. For example, Andergassen, Mödritscher, and Neumann (2014) used log data from a learning management system (LMS) to analyze exam preparation behaviors, finding that practice and repetition were positively correlated with final exam grades. Other researchers have identified behaviors associated with negative learning outcomes, such as cheating (Baker, Corbett, Koedinger, & Wagner, 2004; Cocea, Hershkovits, & Baker, 2009), discerning a student's level of motivation and engagement through the analysis of time stamps (D'Mello, Olney, & Person, 2010; McQuiggan, Mott, & Lester, 2008), and students' affective state and its role in the learning process (D'Mello, Lehman, & Graesser, 2011). Researchers have also investigated how educational data mining and LA might help us understand metacognition and motivation to learn (Pardos et al., 2014; Winne & Baker, 2013).

Theoretical Framework

Learning Analytics in Evaluation and Implementation Research

We propose a new area for growth for the field of LA in implementation and evaluation research. Specifically, educational researchers should leverage LA techniques along with teacher and student data to measure three of the dimensions of fidelity of implementation, summarized in Table 2. First, LA techniques and user data can provide a measure of *adherence*, that is, whether a user is utilizing the digital innovation in the intended way. This can help the researcher uncover patterns in how a user interacts with the program, including the order in which resources were accessed, how long a user remained on a given resource, and with what actions in that resource the user engaged. Second, LA techniques and user data can measure *dosage* or *exposure*, or how frequently and for how long a user engages with the digital innovation. Third, they can measure how *engaged* a user is with the digital innovation by looking at, for example, time spent on a page or component or how quickly a user moves through components. Researchers can learn more about the impact of implementing different components of a digital innovation on student learning by examining how a user interacts with the various components and weighing these data with other dosage data to estimate the impact of different components (Nelson et al., 2012).

It is worth emphasizing that using user data do not (and should not) supplant existing fidelity measures; rather, they can complement them. User data can provide varying levels of information on the interaction between the user and the digital innovation, but they do not capture the quality of that implementation. For this reason, it is important to complement the use of LA techniques and user data with observations, interviews, and other traditional measures of FOI.

Table 2. Dimensions of Fidelity of Implementation and User Data

Dane & Schneider, 1998; Nelson et al., 2012	Hagermoser, Sanetti, & Kratochwill, 2009	Century et al., 2012	User Data
Adherence	Content	FOI of instructional materials	X
Exposure	Quantity	Structural/ procedural	X
Quality of delivery	Quality/process	Instructional/pedagogical	
Participant responsiveness		Student engagement	X

Considerations for utilizing LA techniques and user data. LA techniques and user data can shed light on how digital interventions are used. In order to realize this potential, however, there are several considerations that should be addressed.

First, what are the user data measuring? The answer is contingent on the architecture in place, that is, the way the website or program is set up. Different configurations allow for the collection of different kinds of data, which will be more or less fine-grained. Ideally, the system should collect very fine-grained data that can become rich information about user engagement and learning, but this level of data capture is far from ubiquitous, particularly in K–12 education.

A second consideration is data reliability and validity. Reliability and validity are closely tied to the first question—what is being measured. For example, many websites, including the one evaluated here, collect only "page clicks," and these clicks may not tell us what we think they are telling us. That is, unfiltered page clicks capture a wide range of movement through a website, including accidental clicks, exploratory clicks, and engaged clicks. It is difficult to know how to differentiate among these actions.

Researchers may weigh different options to address these two challenges. One approach is to create a more complex event capture system so that the data collected are more fine-grained. For example, in order to study engineering design processes among high school students, Xie and colleagues implemented a methodology that used computer-aided design (CAD) logs data (Xie, Zhang, Nourian, Pallant & Bailey, 2014; Xie, Zhang, Nourian, Pallant, & Hazzard, 2014). These data correspond to all interactions in a CAD tool. A second approach would be to create filters that set thresholds for when data will be categorized as a particular event, which allows researchers to begin to differentiate among events based on specific parameters. For example, we could determine when a particular instructional component should be considered "viewed" by requiring that a student must watch a certain percentage of a video.

A third solution would be to incorporate metadata (event descriptors) or semantic information into the analytics data. For example, a click on a simulation game can be associated with a particular activity embedded within a larger lesson cycle. Semantic information refers to descriptions about relationships or properties about events. For example, using resource X followed by activity Y can be categorized as scaffolding learning (supporting the student as he or she moves to the next level of instruction), whereas using resource X followed by activity Z may be a learning experience at the same level of instruction. With these more sophisticated data, researchers can, for example, identify different sequences of events within a given period of time that are associated with one specific learning activity. Taken together, a pattern of sequenced events might then represent a broader lesson and have implications for the pedagogy underpinning the lesson and its activities. Researchers also can examine the time elapsed between events and the overall duration of a sequence to better understand the effect of sequencing and pacing on learning outcomes.

A third consideration is access to technology. Technology access can be a challenging barrier to overcome for digital interventions and the barrier remains quite widespread, particularly in low-income communities (Mossberger, Tolbert, & Gilbert, 2006; Warschauer & Matuchniak, 2010). When students do not have access to computers or devices, it is likely that analysts will encounter additional reliability problems in the data collected. For example, if a classroom does not have enough computers, it may be difficult to implement the program as intended (Heffernan, Militello, Heffernan, & Decoteau, 2012).

A fourth consideration is data privacy. Though a full discussion of the ramifications of working with user data and so-called "big data" is beyond the scope of this article, it is worth raising this issue. Improvements both in digital educational programs and in LA techniques mean that developers and researchers have access to unprecedented amounts of user data, both for teachers and students. In most cases, users agree to a program's terms of use without fully understanding what they are agreeing to, subsequently handing over data to developers and researchers (Rowan & Dehlinger, 2014; Solove, 2013). Willis, Campbell, and Pistilli (2013) discuss some ethical considerations for using learning analytics in higher education, given how much we can learn from these data, and many of these considerations apply to K–12 education.

An Example of LA Integration Into Program Evaluation

Our thinking on the utility of learning analytics and user data has emerged from our ongoing work in program evaluation. As part of an evaluation of a digital science curriculum in which teacher user analytics data have been central, we have experimented with different ways to manipulate teacher user data, using techniques from LA, to measure teacher implementation of the curriculum. We used the following research questions regarding the implementation and impact of the curriculum to guide our analyses:

(1) How did teachers use the curriculum?
(2) Was teacher use of the curriculum associated with student performance on the state science test?

Question 1 asks about FOI, while question 2 asks about the associations between FOI and student learning outcomes.

Method

Program Description

The commercially available curriculum under examination, STEMscopes (http://acceleratelearning.com), is a comprehensive online science curriculum that provides hands-on inquiry activities, assessments, problem-based learning (PBL) activities, intervention tools, acceleration materials, and teacher resources for grades K–12, although this study focuses on elementary schools only. The curriculum is meant to be implemented in a "blended" classroom, where some activities are completed online while others are completed face-to-face. In the case of the version of the curriculum we evaluated and report on here, all of the materials were housed online, but most of the activities are completed face-to-face. The majority of interactions with the curriculum logically came from the teachers as they accessed activities, read content background information, or downloaded student worksheets. For this reason, we focused data collection and analysis efforts on teachers' use of the curriculum.

Pedagogical foundations. The curriculum is a variation the 5E model (Bybee et al., 2006; Zuiker & Whitaker, 2014) and offers several activities for each of its steps. The 5E refers to five "steps" of learning science: engagement, exploration, explanation, elaboration, and evaluation. In addition to these five steps, the curriculum adds two steps to help teachers differentiate their instruction according to what their students need, making it the 5E + I/A model. The first of these extra steps, *Intervention*, comprises activities that provide students with additional opportunities to master content and teachers with additional tools to assess whether students have mastered the content after the intervention. The second of these steps, *Acceleration*, provides activities for students who have demonstrated mastery of the content and are ready to extend or apply their learning through short projects, including PBLs and science-related art (Zuiker & Whitaker, 2014).

Critical components. In the curriculum, there is a hierarchy of critical components. Century et al. (2012) define critical components as "the intervention components that developers believe are essential to achieving desired outcomes as a result of innovation implementation" (p. 348). The curriculum's components can be divided into groups according to the 5E + I/A pedagogical model and are summarized in Table 3. In the *Engage* component, there are supports for teachers and activities for students to complete in order to pique students' interest and activate their existing content knowledge. *Explore* centers around a hands-on activity and teacher materials to facilitate these activities. *Explain* provides an opportunity for students to reflect on the Explore activity and to articulate their understanding of the content in their own words.

The remaining steps provide opportunities to extend and evaluate students' learning and to support students who have not mastered the content. In *Elaborate*, there is a third hands-on activity and ideas to extend students' learning. The *Evaluation* step focuses on formal assessment of what students have learned and offers several different kinds of assessments. The final two columns,

Table 3. Critical Components of STEMscopes

Essentials	Engage	Explore	Explain	Elaborate	Evaluate	Intervention	Acceleration
Student expectation	Teacher instructions for demonstration	Teacher guide	Question prompts	Next step inquiry teacher instructions	Pre-assessment	Guided practice	Problem-based learning activity
Key concepts	Starter ideas and activities	Setup video	Picture vocabulary	Next step inquiry student materials	Progress monitoring assessment	CLOZE passages	Art related to science
Fundamental questions	Student materials for Engage activity	Explore student materials (student journal and guide)	Reader	Leveled reading passages on-topic	Standards-based assessment	Assessment	
Teacher background	Science rock		Vocabulary game	Books on topic	Open-ended responses		
Standards correlations			Vocabulary cards	Additional web resources	Performance assessment		
Materials list				Ideas and activities for extensions	Writing science prompts		
Scope summary				Engineering connections	Review game		
Answer keys							
TEKS unwrapped							

Intervention and *Acceleration*, offer additional activities to support students struggling to understand the content and to challenge students who have demonstrated mastery.

Analytics architecture. STEMscopes data sets are updated in real time as teachers and students interact with the curriculum (e.g., click on resources in the website). Therefore, the volume of data stored and processed grows constantly, with an average of a 250,000 daily interactions during the 2012–2013 school year. To tackle this problem, we have implemented a distributed data warehouse environment, employing four separate high-end computers. Anonymized event data generated by users interacting with the curriculum are processed, and then analytics variables are calculated and aggregated at different levels such as by user (teacher or student), school, and district. Currently, our data include information about teacher and student page visits by science standard, pedagogical step (5E+I/A framework), and activity used, but, as mentioned earlier, most components on this version of the curriculum are meant for teacher use, and those meant for students comprised only online games and assessments. Analytics variables are also calculated at the end of each week, month, and year. For a more detailed description of the architecture and ongoing work, see Monroy et al., 2013 and 2014.

Data collection. As part of the program evaluation, we collected FOI data in several ways, including classroom observations, teacher surveys, teacher interviews, and user data. These data were collected in a large, urban school district in Texas. Here we focus on how we utilized LA techniques together with teacher user data to examine how teachers in this district used the curriculum, and what these data revealed about implementation fidelity. We also compare the user data to the survey responses to examine the discrepancy between the two.

We administered a teacher survey in January 2013 and received student and teacher data from the district, including demographics and student science scores from the 2012–2013 school year. The user data consisted of page visits, where one data point is an instance when a teacher accessed a component on the curriculum's webpage (in this study, we did not collect student learning analytics data because of a lack of access to computers in classrooms).

Teacher survey. We administered an online survey to teachers with active STEMscopes accounts in the target school district. The survey, adapted the survey from the *Survey of Educator Data Use* (Wayman, Cho, & Shaw, 2009b), asked teachers about themselves and their classroom, and also measured seven concepts related to science, the curriculum, and technology in the classroom. The scales of interest in this study was the *Use of 5E + I/A* scale (8 items), which asked teachers to report how frequently they used each of the steps from the 5E + I/A model (reliability for this scale was 0.91, as measured by Cronbach's alpha).

In this particular district, only about 30 schools were using the curriculum at the time of the survey because it had not been adopted district-wide. The survey was administered to 755 teachers at schools with curriculum accounts in the district. We selected these teachers randomly by loading the full population of teachers at schools with accounts (approximately 1,500) into the software package, SPSS 20.0, and requesting that it select a random sample of 755, or about half of the full population (we did not use the full population at the request of the district). Of the 755 teachers surveyed, 210 teachers responded, for a response rate of 28% (we tried to increase response rate by offering an incentive and by requesting support from the district; the district declined to provide any assistance). The sample we utilized for these analyses was further refined when we matched

Table 4. Descriptive Statistics for Teachers Completing Survey

Years experience in teaching	0–5 38.50%	6–10 17.90%	11–20 28.20%	20+ years 15.40%
Access to computers in classroom	1–3 52.10%	4–6 40.80%	7–10 5.60%	One for each student 1.4
Years using curriculum	First year 58.4%	Second year 28.6%	Used predecessor 13.0%	
Training on curriculum	None 52.60%	1 hour 26.90%	2 hours 14.10%	3 hours 6.40%

Table 5. Use Indices Combining User and Survey Data

Variable	Definition
Engage Index	Average use of Engage from analytics and survey responses
Explore Index	Average use of Explore from analytics and survey responses
Explain Index	Average use of Explain from analytics and survey responses
Elaborate Index	Average use of Elaborate from analytics and survey responses
Evaluate Index	Average use of Evaluate from analytics and survey responses

teachers' survey responses to their user data. There were several reasons for this, including deleted accounts, and in the end we worked with a smaller sample of teachers who had both survey responses and user data. Table 4 provides descriptive data for the teachers in this sample.

User data. Currently, the architecture supporting the online science curriculum allows only for the tracking of page visits, so all of the data we collected consisted of instances when a teacher accessed a particular page of the curriculum, regardless of what the teacher does or how long the teacher is on the page (Monroy et al., 2013, 2014). This architecture, therefore, imposes important limits on the analyses: We cannot be sure exactly what kind of activity or what level of engagement the data are measuring, so the validity of the data can be questioned. As a result, the teacher user data we used in this analysis measured dosage—frequency of access to the curriculum—and did not allow us to measure adherence or other dimensions of FOI. More robust data analytics frameworks would allow researchers to measure the dimensions discussed earlier. The limits imposed by the user data mean that it is even more important to triangulate—to combine the user data together with other, more traditional, measures of utilization or implementation. The work presented here represents an initial step toward more sophisticated analyses and still provide lessons for researchers doing similar work.

Though we had access to data for every single activity within the curriculum, we decided to focus only on the critical components. We also created a set of variables that combined the user and survey data for each teacher (see Table 5). These variables were meant to help us triangulate—to bring together two different measures of the same behavior to try to paint a more complex picture of teacher utilization. We recognize that creating a detailed, valid, and reliable measure of instruction is more complex than what we are able to accomplish here, but this is one step toward creating new measures.

District data. District data were collected including demographic data for fifth-grade science teachers, and demographic and achievement data for fifth-grade students for the 2012–2013 school year. For students, data included sex, race/ethnicity, socioeconomic status, and certain academic data (e.g., limited English proficiency status, special education status, and gifted and talented status) because the literature on student achievement suggests they all are important demographic

Table 6. Descriptive Statistics of for Fifth-Grade Teachers and Students Matched to District Data

Variable	STEMscopes Schools
Fifth-grade teachers	$n = 18$
Average teacher experience	12.2 years
Average number of students	48.6
New teachers	38%
Fifth-grade students	$n = 916$
Female	48%
White	3%
African American	14%
Hispanic	82%
Asian	1%
Multiracial	<1%
Other race/ethnicity	<1%
Free/reduced lunch	57%
Limited Engish proficiency (LEP)	57%
Bilingual (BIL)	31%
English as a second language (ESL)	1%
Special education	9%
Gifted/talented	23%

predictors. For teachers, data included whether a teacher was new to the profession and years of teaching experience, both of which have been shown to help explain teacher effectiveness. Once we matched data from the survey and the teacher user data, we ended up with 18 fifth-grade teachers. Of these 18 fifth-grade teachers, 38% of them were new, and the average number of years of experience was 12.2 years (see Table 6 or complete demographic data).

Data analysis. To answer research question 1, we describe teacher use of the curriculum as measured by the survey and user data in order to give an overview of how the picture of use varied across the two measures.

For research question 2, we utilized hierarchical linear modeling (HLM) to estimate the association between teacher use of STEMscopes on students' scores on the state science achievement test (STAAR) given to fifth-grade students at the end of the 2012–2013 school year. HLM allows the researcher to control for possible confounding factors at each level of analysis (e.g., gender, ethnicity, etc.) and to take into account the nested structure of the data (Raudenbush & Bryk, 2002). We estimated a two-level HLM model with students at Level 1 and teachers at Level 2. Level 1 predictors were centered at the group mean in order to create Level 1 coefficients that were independent from Level 2 variance, and Level 2 predictors were centered at the grand mean to create Level 2 coefficients centered around the average score of all schools for interpretability (Enders & Tofighi, 2007; Raudenbush & Bryk, 2002).

We experimented with the different variables as ways to measure dosage. First, we analyzed the student test scores utilizing only the teachers' survey responses related to their use of the 5E model. Then, we conducted the same analyses using only the teachers' user data describing their access of the 5E model. Finally, we created indices that combined the teachers' survey responses and their user data as a way to triangulate the two. The equation for the final multilevel models can be seen in Equation 1:

$$\begin{aligned} Level1: STAAR\ Science\ Score_{ij} &= \beta_{0j} + \beta_{1j}(Sex) + \beta_{2j}(\text{Race/Ethnicity}) \\ &+ \beta_{3j}(\text{Free/Reduced Lunch Status}) + \beta_{4j}(LEP) + \beta_{5j}(\text{Bilingual}) \\ &+ \beta_{6j}(ESL) + \beta_{7j}(\text{Special Education}) + \beta_{8j}(\text{Gifted/Talented}) + r_{ij} \\ Level2: \beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{Teacher Use}) + u_{0j} \end{aligned} \tag{1}$$

Results

Research Question 1

Research question 1 asked how teachers used STEMscopes in the classroom. We explored this question using various types of data and here we begin by discussing trends in teacher use that the survey and user data revealed.

Teacher survey. The survey asked teachers how often they used the components of the curriculum. The survey items gave us information on dosage, or frequency of use, and on adherence, since one of the items also asks teachers how often they use the steps in order. The survey results revealed "medium" use of the curriculum and some of its core components (see Table 7). The two most used steps were Explain and Evaluate, though use of Explore and Engage was not too much lower.

Table 7. How Often Do You Use the Following Categories in Your Lessons?

Lesson Category	Every TEKS	Most TEKS	Some TEKS	No TEKS
Engage	17%	38%	36%	9%
Explore	19%	38%	35%	8%
Explain	20%	33%	39%	8%
Elaborate	13%	28%	46%	13%
Evaluate	24%	30%	36%	10%
Intervention	7%	15%	60%	18%
Accelerate	4%	9%	57%	30%
In order?	22%	33%	37%	8%

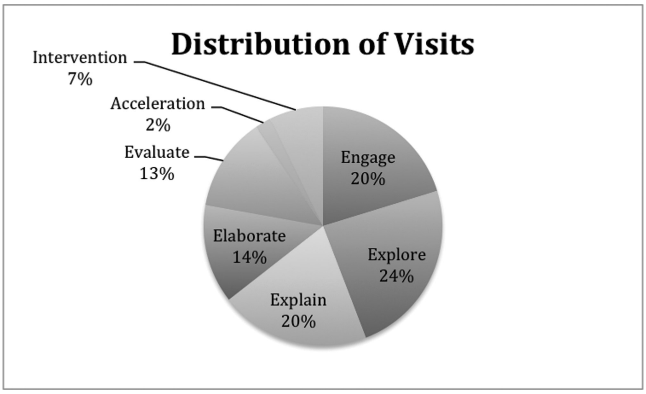

Figure 1. Distribution of visits across teachers, by 5E + I/A step.

Teachers reported rarely using the two additional steps, Intervention and Accelerate. In the next section, we present descriptive results of teacher use derived from the user data collected for teachers with active STEMscopes accounts in the district.

Teacher user data. We used the teacher user data to measure dosage and adherence. The level and kind of use that emerged from the user analytics data are similar to what we learned from the teacher survey. When examining aggregate teacher use of the 5E + I/A steps (see Figure 1), the most frequently visited steps were Explore, Engage, and Explain. The least used steps were Intervention and Acceleration. One clear difference is that the Evaluate step was not the most visited step, and yet, according to the survey, Evaluate was the step that teachers used most frequently.

Research Question 2

In research question 2, we examined how different measures of teacher use were associated with student state science scores, measured by the STAAR exam, controlling for important student demographic covariates. The first model we estimated examined associations between teacher survey responses about their use of the primary 5E steps and student STAAR scores. We included their responses to all five items in the model and found that only their reported use of the Engage step was significantly related to student scores, and this relationship was negative: The more teachers reported using the Engage step, the lower their students' scores were (see Table 8).

Table 8. Associations Between Reported Use and Student Science Scores

	STAAR Science Scores		
Predictor	*B*	*SE*	*p* Value
Teacher-level predictors			
Engage	–174.42*	74.42	.039
Explore	130.76	60.56	.054
Explain	–27.22	42.09	.531
Elaborate	27.43	56.33	.636
Evaluate	88.97	45.54	.077
Student-level predictors			
Sex	–76.62*	23.86	.001
Race/ethnicity	116.15***	27.80	<.001
Free or reduced lunch	–2.12	3.47	.542
Limited Engish proficiency (LEP)	22.82	16.87	.177
Bilingual (BIL)	–53.12***	7.36	<.001
English as a second language (ESL)	–204.25	149.37	.172
Special education	–395.64***	42.83	<.001
Gifted/talented	361.64***	32.47	<.001

Note. Science scores represent student scaled scores (Range = 1,000–6,000).
*** $p < .001$. ** $p < .01$. * $p < .05$.

Table 9. Associations Between Teacher Analytics Data and Student Science Scores

	STAAR Science Scores		
Predictor	*B*	*SE*	*p* Value
Teacher-level predictors			
Engage	–1.50**	0.47	.009
Explore	0.36*	0.13	.021
Explain	0.17	0.92	.860
Elaborate	1.10*	0.38	.015
Evaluate	0.95	0.52	.092
Student-level predictors			
Sex	–76.63*	23.84	.001
Race/ethnicity	105.45***	27.07	<.001
Free or reduced lunch	–2.13	3.47	.539
Limited Engish proficiency (LEP)	21.48	16.76	.200
Bilingual (BIL)	–51.29***	7.34	<.001
English as a second language (ESL)	–181.00	148.79	.224
Special education	–397.90***	42.73	<.001
Gifted/talented	368.21***	32.48	<.001

Note. Science scores represent student scaled scores (Range = 1,000–6,000).
*** $p < .001$. ** $p < .01$. * $p < .05$.

We then estimated a model that incorporated the teacher user data, which produced similar results. Teacher access of the Engage step continued to be negatively related to student science scores. We also found, however, that teacher access of the Explore and Elaborate steps was positively related to student scores, but the effects were smaller (see Table 9).

Finally, we estimated a model using the indices we created combining the survey data and the user analytics data (see Table 10). Three of the 5E indices were related to student scores. Teacher use of the Engage step, as measured by teacher survey responses and user data, continued to be negatively associated with student science scores. However, teacher use of the Explore step and teacher use of the Elaborate step were positively associated with student science scores. To summarize, teachers' reported use of the Engage step was negatively related to student achievement; however, when we examined teacher use as measured by the user data and then as measured by our indices, we observed different results: In addition to a significant, negative relationship between teacher use of the Engage step, we also found two positive relationships, between teacher use of Explore and Elaborate and student achievement.

Table 10. Associations the 5E Indices and Student Science Scores

	STAAR Science Scores		
Predictor	*B*	*SE*	*p* Value
Teacher-level predictors			
Engage Index	−3.00**	0.95	.009
Explore Index	0.72*	0.27	.020
Explain Index	0.34	1.84	.858
Elaborate Index	2.19*	0.76	.016
Evaluate Index	1.89	1.02	.092
Student-level predictors			
Sex	−76.63**	23.84	.001
Race/ethnicity	105.43***	27.07	<.001
Free or reduced lunch	−2.13	3.47	.539
Limited Engish proficiency (LEP)	21.47	16.76	.201
Bilingual (BIL)	−51.28***	7.34	<.001
English as a second language (ESL)	−180.97	148.78	.224
Special education	−397.93***	42.72	<.001
Gifted/talented	368.24***	32.48	<.001

Note. Science scores represent student scaled scores (Range = 1,000–6,000).
*** $p < .001$. ** $p < .01$. * $p < .05$.

Discussion

The primary purpose of this article was to discuss how user data and LA techniques can be incorporated into educational research, particularly implementation research. In this way, our work here responds to Hulleman and Cordray's (2009) recommendation that researchers find ways to combine different implementation measures to better understand the impact of utilization. Together, user data and learning analytics can contribute to a more nuanced understanding of implementation, and of the impact of implementation on learning outcomes. We believe user data and learning analytics will become indispensable tools for educational researchers as digital interventions become more widespread and budgets become tighter, requiring educators to vet the efficacy of digital interventions very carefully.

The second purpose was to demonstrate one way these data can be incorporated to examine fidelity of implementation. For this reason, we were less interested in the actual results of the analysis than in the story that the data told about the extent and impact of teacher implementation of the curriculum. Importantly, the survey results and the results of the user data analysis were not identical, which suggests either that they are not, in fact, measuring the same underlying behavior (i.e., teacher implementation of the curriculum), or that they measure differences that stem from the perspective from which the data capture implementation. In other words, the survey results represent self-report data, while the user data represent a digital footprint of sorts (Snodgrass Rangel, Monroy, Bell, & Whitaker, 2013).

Future research should include additional traditional measures of FOI, such as observations, interviews, use logs, and so on. Future work also will have to contend with the considerations relating to what is being measured, data reliability and validity, and technology access. In our analyses, we utilized user data and teacher survey data as complementary measures of dosage and adherence, but were not able to utilize user data to capture other dimensions of FOI. Researchers who decide to incorporate user data into their evaluations and other analyses of digital interventions similarly should identify the limits of the data to which they have access and try to move past that challenge by working with technology developers and school districts to improve their analytics infrastructure.

The second consideration, data reliability and validity, also is key. Our research team was in agreement that our data were limited because they only measure page visits: One cannot know for how long a user was on a page (e.g., did the teacher reach a page by accident?), what a user did on a page (did he or she scroll up and down? right-click?), or whether something on the page was downloaded and printed for later use. In other words, the more detailed the data, the more likely it is that they will yield valid and reliable insights. What's more, researchers should conduct analyses to confirm the reliability or validity of the data to ensure the quality of the variables created.

Finally, our research finds additional support for work that points to the persistent limited access to technology in many public schools; access to computers was a challenge in the district where this evaluation was conducted. The impact of this divide for researchers is that if teachers and students do not have regular and reliable access to computers or other digital devices, teachers and students may not use the program as intended (or at all), and the user data are unlikely to be valid or reliable. Efforts to ensure access to technology as well as support and training for appropriate technology use should be a central focus for researchers and developers of digital instructional resources. As these efforts make progress, incorporating learning analytics and user data into educational research will become more critical, and this work can serve as a road map for future work.

Received: 11/8/13
Initial decision: 10/22/14
Revised manuscript accepted: 12/5/14

Acknowledgments. We thank the anonymous reviewers for their thoughtful reviews, which greatly improved our article.

Declaration of Conflicting Interests. During the time this research was conducted, the first three authors were researchers with the Center for Digital Learning and Scholarship, which helped to develop the STEMscopes curriculum. J. Reid Whitaker is currently the Chief Academic Officer for Accelerate Learning Inc., which owns the STEMscopes curriculum (since 2013), and was the executive director of the Center for Digital Learning and Scholarship at Rice University.

Funding. The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Materials. A supplemental appendix for this article can be accessed on the publisher's website at http://dx/doi.org/10.1080/15391523.2015.999639

Author Notes

Virginia Snodgrass Rangel is an independent education researcher and owner of Third Coast Insight in Houston, Texas. Her research interests focus on program evaluation and implementation research, particularly in science education and educational technology and data use. Please address correspondence regarding this article to Virginia Snodgrass Rangel, 662 Lester Street, Houston, TX 77007, USA. E-mail: virginia.srangel@gmail.com

Elizabeth R. Bell is research manager at Accelerate Learning Inc. Her research interests focus on early science education and program evaluation.

Carlos Monroy is research scientist in the Department of Computer Science at Rice University. His research interests focus on learning analytics, data mining, and information visualization.

J. Reid Whitaker is currently the chief academic officer for Accelerate Learning Inc. His research interests include leadership, science education, and blended learning.

References

Andergassen, M., Mödritscher, F., & Neumann, G. (2014). Practice and repetition during exam preparation in blended learning courses: Correlations with learning results. *Journal of Learning Analytics, 1*(1), 48–74.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students game the system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383–390). New York, NY: ACM.

Berman, P., & McLaughlin, M. W. (1976, March). Implementation of educational innovation. *The Educational Forum*, *40*(3), 345–370. http://dx.doi.org/10.1080/00131727609336469

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief.* Washington, DC: SRI International.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). *The BSCS 5E instructional model: Origins and effectiveness*. Colorado Springs, CO: BSCS.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218. http://dx.doi.org/10.1177/1098214010366173

Century, J., Cassata, A., Rudnick, M., & Freeman, C. (2012). Measuring enactment of innovations and the factors that affect implementation and sustainability: Moving toward common language and shared conceptual understanding. *The Journal of Behavioral Services and Research*, *39*(4), 343–361. http://dx.doi.org/10.1007/s11414-012-9287-x

Chumley-Jones, H. S., Dobbie, A., & Alford, C. L. (2002). Web-based learning: Sound educational method or hype? A review of the evaluation literature. *Academic Medicine*, *77*(10), S86–S93. http://dx.doi.org/10.1097/00001888-200210001-00028

Cocea, M., Hershkovitz, A., & Baker, R. (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate? In V. Dimitrova, R. Mizoguchi, B. Du Boulay, & A. Graesser (Eds.), *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building learning systems that care: From knowledge representation to affective modelling. Frontiers in artificial intelligence and applications* (pp. 507–514). Amsterdam, the Netherlands: IOS Press.

Corbi, A., & Burgos, D. (2014). Review of current student-monitoring techniques used in elearning-focused recommender systems and learning analytics. The Experience API & LIME model case study. *International Journal of Artificial Intelligence and Interactive Multimedia*, *2*(7), 44–52. http://dx.doi.org/10.9781/ijimai.2014.276

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–45. http://dx.doi.org/10.1016/S0272-7358(97)00043-3

Dietz-Uhler, B., & Hurn, J. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, *12*(1), 17–26.

D'Mello, S. K., Lehman, B., & Graesser, A.C. (2011). A motivationally supportive affect-sensitive autotutor. In R.A. Calvo & S.K. D'Mello (Eds.), *New perspectives on affect and learning technologies*, (pp.113–126). New York, NY: Springer.

D'Mello, S., Olney, A., & Person, N. (2010). Mining collaborative patterns in tutorial dialogues. *Journal of Educational Data Mining*, *2*(1), 2–37.

Doan, T. A., Zhang, J., Tjhi, W. C., & Lee, B. S. (2011). Analyzing students' usage of e-learning systems in the cloud for course management. In F.-Y. Yu, T. Hirashima, T. Supnithi, & G. Biswas (Eds.), *Proceedings of the 19th International Conference on Computers in Education: ICCE 2011* (pp. 297–301). Bangkok, Thailand: National Electronics and Computer Technology Center.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3-4), 327–350. http://dx.doi.org/10.1007/s10464-008-9165-0

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138. http://dx.doi.org/10.1037/1082-989X.12.2.121

Franks, R. P. & Schroeder, J. (2013). Implementation science: What do we know and where do we go from here? In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems*, (pp.5–20). Baltimore, MD: Brooks.

Fullan, M., & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research*, *47*, 335–397. http://dx.doi.org/10.3102/00346543047002335

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review*, *38*(4), 445–459.

Heffernan, N., Militello, M, Heffernan, C., & Decoteau, M. (2012). Effective and meaningful use of educational technology: Three cases from the classroom. In C. Dede & J. Richards (Eds.), *Digital teaching platforms* (pp. 88–102). New York, NY: Teachers College Press.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, *2*(1), 88–110. http://dx.doi.org/10.1080/19345740802539325

Khribi, M. K., Jemni, M., & Nasraoui, O. (2015). Recommendation systems for personalized technology-enhanced learning. In Kinshuk & R. Huang (Eds.), *Ubiquitous learning environments and technologies* (pp. 159–180). Berlin, Germany: Springer.

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research*, *43*(4), 489–510. http://dx.doi.org/10.2190/EC.43.4.d

Loftus, W. (2012) Demonstrating success: Web analytics and continuous improvement. *Journal of Web Librarianship*, *6*, 45–55. http://dx.doi.org/10.1080/19322909.2012.651416

McLaughlin, M. (1976). Implementation as mutual adaptation: Change in classroom organization. *Teachers College Record*, *77*(3), 339–351.

McQuiggan, S. W., Mott, B. W., & Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-adapted Interaction*, *18*(1–2), 81–123. http://dx.doi.org/10.1007/s11257-007-9040-y

Means, B., Anderson, K., & Thomas, S. (2012). *Expanding evidence approaches for learning in a digital world*. Washington, DC: U.S. Department of Education.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC: U.S. Department of Education.

Mendicino, M., Razzaq, L., & Heffernan, N. T. (2009). A comparison of traditional homework to computer-supported homework. *Journal of Research on Technology in Education*, *41*(3), 331–359. http://dx.doi.org/10.1080/15391523.2009.10782534

Monroy, C., Rangel, V. S., & Whitaker, R. (2013). STEMscopes: Contextualizing learning analytics in a K–12 science curriculum. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 210–219). New York, NY: ACM Press.

Monroy, C., Rangel, V. S., & Whitaker, R. (2014). A strategy for incorporating learning analytics into the design and evaluation of a K–12 science curriculum. *Journal of Learning Analytics, 1*(2), 94–125.

Mossberger, K., Tolbert, C., & Gilbert, M. 2006. Race, place and information technology. *Urban Affairs Review*, *41*(4): 1–38.

Muñoz-Merino, P. J., Valiente, J. A. R., & Kloos, C. D. (2013, April). Inferring higher level learning information from low level data for the Khan Academy platform. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 112–116).

Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services and Research*, *39*(4), 374–396. http://dx.doi.org/10.1007/s11414-012-9295-x

Olmos, M., & Corrin, L. (2012). Learning analytics: A case study of the process of design of visualizations. *Journal of Asynchronous Learning Networks*, *16*(3), 39–49.

Özyurt, Ö., Özyurt, H., Baki, A., & Güven, B. (2013). Integration into mathematics classrooms of an adaptive and intelligent individualized e-learning environment: Implementation and evaluation of UZWEBMAT. *Computers in Human Behavior*, *29*(3), 726–738. http://dx.doi.org/10.1016/j.chb.2012.11.013

Pardos, Z. A., Baker, R. S. J. D., San Pedro, M. O. C. Z, Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*. UTS ePress, *1*(1), 107–128.

Penuel, W. R., & Means, B. (2004). Implementation variation and fidelity in an inquiry science program: Analysis of GLOBE data reporting patterns. *Journal of Research in Science Teaching*, *41*(3), 294–315. http://dx.doi.org/10.1002/tea.20002

Quint, J. C., Balu, R., DeLaurentis, M., Rappaport, S., Smith, T. J., & Zhu, P. (2013). *Early findings from the Investing in Innovation (i3) scale-up*. New York, NY: MDRC.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: CA: Sage.

Rowan, M., & Dehlinger, J. (2014). Privacy incongruity: An analysis of a survey of mobile end-users. In *Proceedings of the 13th International Conference on Security and Management*. Detroit, MI: University of Detroit Mercy.

Scheirer, M. A., & Rezmovic, E. L. (1983). Measuring the degree of program implementation. *Evaluation Review*, *7*(5), 599–633. http://dx.doi.org/10.1177/0193841X8300700502

Siemens, G. (2012, April). Learning analytics: Envisioning a research discipline and a domain of practice. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 4–8). New York, NY: ACM.

Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, *16*(3), 51–61.

Snodgrass Rangel, V., Bell, E. R., Monroy, C., & Whitaker, J. R. (2013, April). *A year in review: Year 1 results from an evaluation of an online blended science curriculum model in an urban district*. Poster presented at the American Educational Research Association Annual Meeting, San Francisco, CA.

Solove, D. J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review*, *126*, 1880.

Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the What Works Clearinghouse. *Educational Evaluation and Policy Analysis*, *32*(2), 351–71. http://dx.doi.org/10.3102/0162373710373389

Turner, S. J. (2010). Website statistics 2.0: Using Google Analytics to measure library website effectiveness. *Technical Services Quarterly*, *27*(3), 261–78. http://dx.doi.org/10.1080/07317131003765910

Wang, H., & Woodworth, K. (2011, September). A randomized controlled trial of two online mathematics curricula. Paper presented at the Society for Research on Educational Effectiveness Fall 2011 Conference, Washington, DC.

Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*, *34*(1), 179–225. http://dx.doi.org/10.3102/0091732X09349791

Wayman, J. C., Cho, V., & Shaw, S. M. (2009a). *First-year results from an efficacy study of the Acuity data system*. Austin, TX: University of Texas.

Wayman, J. C., Cho, V., & Shaw, S. (2009b). *Survey of educator data use*. Austin, TX: University of Texas at Austin.

Wayman, J. C., Shaw, S. M., & Cho, V. (2011). *Second-year results from an efficacy study of the Acuity data system*. Austin, TX: University of Texas at Austin.

Willis, J. E. III, Campbell, J., & Pistilli, M. (2013). Ethics, big data, and analytics: A model for application. *Educause Review Online.* Retrieved from http://apo.org.au/research/ethics-big-data-and-analytics-model-application

Winne, P. H., & Baker, R. S. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 1–8.

Xie, C., Zhang, H., 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, H., 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.

Yu, Taeho & Jo, Il-Hyun. (2014). Educational technology approach toward learning analytics: Relationship between student online behaviour and learning performance in higher education. *Proceedings of the Fourth International Conference on Learning*, pp. 269–270. http://dx.doi.org/10.1145/2567574.2567594

Zuiker, S., & Whitaker, J. R. (2014). Refining inquiry with multi-form assessment: Formative and summative assessment functions for flexible inquiry. *International Journal of Science Education*, *36*(6), 1037–1059. http://dx.doi.org/10.1080/09500693.2013.834489