

Using Analytics for Improving Implementation Fidelity in an Large Scale Efficacy Trial

Mingyu Feng, Jeremy Roschelle, Robert Murphy
SRI International, 333 Ravenswood Ave, Menlo Park CA 94025, USA
Email: mingyu.feng@sri.com, jeremy.roschelle@sri.com, robert.murphy@sri.com

Neil T. Heffernan, Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609
nth@wpi.edu

Abstract: The field of learning analytics is rapidly developing techniques for using data captured during online learning. In this article, we develop an additional application: the use of analytics for improving implementation fidelity in a randomized controlled efficacy trial. In an efficacy trial, the goal is to determine whether an innovation has a beneficial effect under best-case implementations. Analytics is more accurate and less expensive than traditional ways of collecting and analyzing implementation fidelity data, and may allow targeted adaptations of the innovation that improve the quality of the research. We report our experience in developing and using analytics during the course of an efficacy trial that evaluated the use of ASSISTments as an online homework tool for middle school mathematics.

Significance

The fields of learning analytics and educational data mining are rapidly developing new techniques for using the copious data that is captured during online learning. Applications of learning analytics have included predicting student outcomes, improving learning resources, and intervening for particular students to enhance their learning trajectories. In this article, we develop an additional application: the use of analytics as a technique for improving implementation fidelity in a randomized controlled efficacy trial.

In an efficacy trial, the goal is to determine whether an innovation has a beneficial effect under best-case implementations. An important contrast is to an effectiveness trial, which aims to measure effects when the innovation is in broader use, with more environmental variation and less control over implementation. Because of the emphasis on best-case interventions in an efficacy trial, it is fair game in an efficacy trial to monitor and adjust implementation of the innovation. Analytics, we will argue, can provide an important new tool for monitoring implementation fidelity, and thus can allow targeted adaptations of the innovation that improve the quality of the research.

Conducting a Randomized Controlled Trial (RCT) is the primary methodology for educational efficacy research. In its basic form, an RCT randomly assigns participants to alternative conditions, where the conditions are deliberately designed to emphasize a desired contrast. Outcomes are measured, and if there is a difference in outcomes in the contrasting conditions, then the contrasting features of the two conditions are presumed to cause the difference. This inferential process depends on the quality of contrast as experienced by the participants: if the contrast is weak, or highly variable, or drifts away from the intended contrast, then measured effects may be due to something other than the designed contrast. Thus, it is important to understand the contrast between conditions as implemented, which traditionally leads to the idea of implementation fidelity—are the conditions implemented in a way that highlights the planned difference between conditions? Is the treatment condition being implemented in a way that preserves the potential for a beneficial effect?

When an efficacy trial is conducted in schools, collecting and analyzing implementation fidelity data is typically slow, inaccurate, and/or expensive. Indeed, often the analysis of implementation fidelity only occurs after the experiment is complete—which can be wasteful if it turns out that the desired contrast was not implemented well, and therefore the experiment is invalid (i.e. the investigators can obtain “no effect” because the treatment was not implemented well according to the model specified by innovation developers, not because the treatment could not produce benefits). Traditional methods for collecting implementation fidelity data are through observations or through self-report. Observations are expensive to conduct, and in contexts where 50 or 100 schools participate in a study, which is usually the case for an efficacy trial, it is typical that projects can only afford one or two observations per school year. Self-report is less expensive to collect, but can be inaccurate. Retrospective interviews are a third source of data, but also introduce concerns about inaccuracy or biases. We argue that analytics are an additional method for collecting implementation fidelity data that can be faster, cheaper, and more objective. In support of this claim, we report our experience in developing and using analytics during the course of an RCT that evaluated the use of ASSISTments as an online homework tool for middle school mathematics in the state of Maine.

Methodological Approach

Implementation Fidelity

Implementation Fidelity is the extent to which the delivery of an intervention conforms to the protocol or program model as intended by the developers of the intervention (Domitrovich & Greenberg, 2000; Mowbray et al., 2003). The assessment of implementation fidelity has been highlighted as critical to understanding how programs are implemented in efficacy studies (Domitrovich & Greenberg, 2000; Durlak & DuPre, 2008; Dusenbury et al., 2003). Despite the value of measuring implementation fidelity in conducting and interpreting RCTs, a long-standing problem is that implementation fidelity is often not measured or underreported, likely due to the expense and difficulty of collecting relevant data.

There have also been critiques of the construct of implementation fidelity: it seems to assume that an innovation should be delivered in the same way in every school, and may better suit over-scripted approaches than highly adaptive approaches. What if implementing a particular innovation “well” means engaging in extensive adaptations of the innovation? As our case study of ASSISTments will show, this concern can be addressed: ASSISTments is intended to be highly adaptive and yet it still makes sense to monitor fidelity, for example, by monitoring whether teachers are using the ASSISTments facility for adapting homework problem sets. Thus it seems possible to develop analytics that detect adaptive or non-adaptive behavior by teachers, and such detectors can contribute to understanding of whether the expected adaptations are likely to be occurring with the innovation.

The literature favors *model-based* approaches, which consider implementation fidelity relative to a logic model (Nelson et al. 2010). A typical logic model traces the causal pathway(s) from affordances designed on the basis of learning theory, to inputs provided to a school (such as new software or teacher professional development), to activities enacted in the school using the inputs, to outcomes that are measured. A sound logic model is central to any high quality efficacy trial.

When measuring implementation fidelity relative to a model, five types of implementation information may be helpful (Cordray, 2008): adherence, exposure, quality of delivery, participant responsiveness, and program differentiation (Durlak & DuPre, 2008; Dusenbury et al., 2003; Fagan et al., 2008). A first pair of implementation measures addresses availability and use of inputs: *Adherence* tracks whether the expected inputs are actually in use at the target schools: do participants access and use the resources provided? *Exposure* monitors how much of the resource is used: is the full extent of the resource used? Are the frequency and dosage of use as intense as the developer recommends? A second pair of implementation metrics addresses the quality of the ensuing activities at schools. *Quality of delivery* reflects the manner in which a program is delivered and can capture whether the activities using the resources are unfolding according to the expected teaching and learning processes. For example, if software attempts to give students practice using the “spacing effect” as recommended by Pashler et al. (2007), are students actually practicing the same skills at regularly spaced intervals? *Participant responsiveness* can look at uptake by teachers and students of the features of the innovation: for example, if the system provides teachers with reports, do they open them? If students have opportunities to choose more challenging problems or to watch tutorial videos, do they do this? Finally, a last category concerns the intended contrast. *Program differentiation* looks at whether the treatment conditions are different from other conditions in expected ways, including mediating processes. For example, if an innovation is expected to increase overall learning by providing more feedback to learners, are we sure that learners in the control condition are not getting the same levels of feedback, but through different processes? We argue that analytics could be developed for these categories of implementation fidelity.

Learning Analytics

The field of educational data mining and learning analytics (LA) has developed rapidly recently (Baker & Yacef, 2009; Romero & Ventura, 2007, 2010; Siemens & Baker, 2012; U.S. Department of Education, 2012). The 2013 Horizon Report (EDUCAUSE, 2013a) describes learning analytics as the “...field associated with deciphering trends and patterns from educational big data, or huge sets of student-related data, to further the advancement of a personalized, supportive system of higher education.” However, LA is not limited to higher education. With technology usage become popular and more accessible among younger children, there has been growing use of LA in K-12 settings (EDUCAUSE, 2013b). The main purpose of LA has been to observe and understand learning behaviors in order to enable appropriate interventions at the individual, course, department, or even institution level (Brown, 2011).

Online learning systems—learning management systems, learning platforms, and learning software—have the ability to capture streams of fine-grained learner behaviors. Then it is the responsibility of data analysts to operate on the data, through procedures such as raw data processing, data aggregation, and/or data modeling using data mining algorithms, in order to make necessary inferences. Different from pure data mining, the process of LA often draws on a broader array of academic disciplines, incorporating concepts and techniques from information science and sociology, in addition to computer science, statistics, psychology, and learning

sciences. A good understanding of the entire learning system and educational environment where the system was used is also needed to draw useful and valid conclusions. Once the data analysis is completed, the findings are provided to a variety of stakeholders who can use the feedback to improve instruction, or improve the learning systems, or other educational decision-making for learners. Thus, the feedback loop is closed.

So far, improving student's learning outcomes has been the core goal of LA. While in this paper, the use of LA supports the closure of a different feedback loop that involves innovation developers, evaluators, and implementers, and implementation supporters.

Case Study: The ASSISTments Efficacy Trial

ASSISTments System and Research Design

ASSISTments (www.assistments.org) is an online tutoring system that provides “formative assessments that assist.” Teachers choose (or add) homework items in ASSISTments and students can complete their homework items online. As students do homework in ASSISTments, they receive feedback on the correctness of their answers. Some students may choose to do their homework offline, but in typical use, teachers still require students to upload their answers before coming to class. Some problem types also provide hints on how to improve their answers, or help decompose multistep problems into parts (see Figure 1). Teachers may choose to assign problem sets called “skill builders” that are organized to promote mastery learning (Anderson, 2000). Teachers also receive reports on their students' homework and can use this information to organize more targeted homework reviews, to assign specific follow-up work to particular students, and to more generally adapt or differentiate their teaching. ASSISTments is provided to schools as a free service of Worcester Polytechnic Institute (WPI). Prior research has found that analytics based on students' usage of the system during the year can predict end-of-year scores on statewide standardized test (Feng et al. 2009; Pardos et al. 2013), identify students engagement states (San Pedro et al., 2013) and college attendance (San Pedro et al., 2013b)..

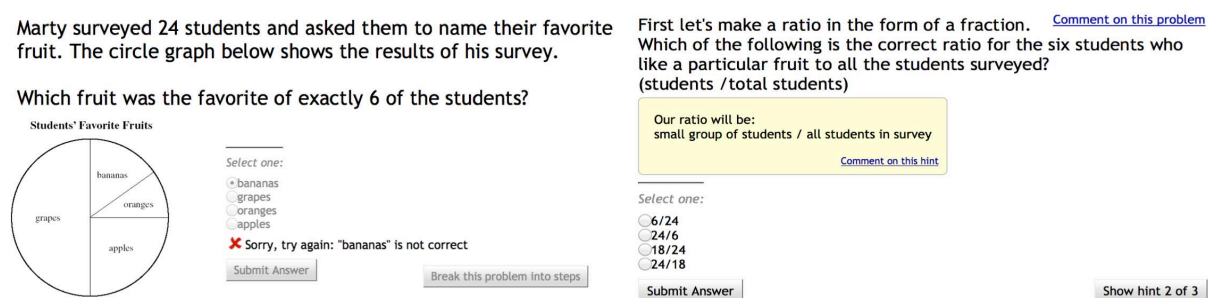


Figure 1. Screen shots of an 7th grade item in ASSISTments that provides correctness feedback and breaks the problem into steps (left) and the first sub-step with a hint message (right)

Prior research also has established the promise of ASSISTments for improving student outcomes in middle school mathematics through homework support (Mendicino, Razzaq, and Heffernan, 2009; Singh et. al, 2011). Building on this prior research, a team led by SRI International in collaboration with WPI and the University of Maine is conducting a large-scale efficacy trial with ASSISTments in the state of Maine where a one-to-one laptop program was well established. The research is an RCT involving 45 middle school schools from two cohorts, with schools randomly assigned to treatment or control (i.e. “business as usual”) conditions. The intervention is implemented in Grade 7 math classrooms in treatment schools over 2 consecutive years (academic years 2012–13 and 2013–14 for Cohort 1 schools and 2013–14 and 2014–15 for Cohort 2 schools). In the Treatment condition, teachers receive professional development (PD) and use ASSISTments in the first year to become proficient with the system and then teachers use ASSISTments with a new cohort of students in the second year when student outcomes are measured. Note that we are testing students in *teachers'* second year of experience with the system, because of the developer's belief that teachers do not sufficiently master the system in their first year of experience.

This design provides a strong opportunity for using analytics for implementation fidelity. Since the goal in the first year is to achieve teacher proficiency with the system, if analytics can reveal whether or not this is occurring in a timely manner, additional mentoring could be provided to bring all teachers up to desired levels of implementation. This can occur *before* the second, measurement year begins.

ASSISTments Logic Model

The efficacy trial is guided by the ASSISTments logic model (See Figure 2). Note that the logic model allows for three pathways to increased student learning. A first path is that students may complete homework

with greater regularity when it is online. Even if there was nothing different about homework online or offline, completing more homework could improve student learning. A second path, labeled “direct effect” is the effect on students of getting support for doing homework. A third path is through reporting to teachers, who can then adapt instruction to their students’ needs. Our strategy was to align potential implementation fidelity analytics to the “features” and “mediating variables” columns of the logic model.

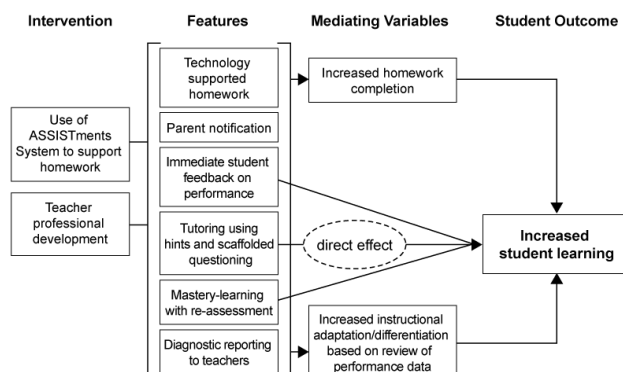


Figure 2. The ASSISTments logic model

Specified Use Model

Based on developers’ prior experience in school implementation, we set the specified use model such that teachers who use ASSISTments in the study are expected to assign approximately 25 minutes of homework in ASSISTments for a minimum of 3 nights per week. Homework assignments created by teachers within ASSISTments are expected to consist of: (1) mastery learning problem sets (aka. “skill builders”) that addresses a prerequisite or recently-instructed mathematics skill; (2) reassessment mastery problems that are automatically assigned by the system and address a skill that a student has previously mastered; (4) and a series of textbook problems that will comprise the majority of the assignment.

Teachers will receive a performance report early the next morning via email, in addition to other reports that they can access after logging into their ASSISTments account. The report informs teachers whether a student completed the assignment, student’s performance on each problem/skill, and also identifies the problems/skills with which most students struggled. Teachers are expected to review (“open”) the homework performance report for a minimum of 50% of assignments.

Design of Candidate Analytics

Our data analytics for the first implementation year, as reported in this paper, center on guiding PD and mentoring offered to the teachers in the first year. Later on, when data has been collected for the overall RCT, the same analytics may be useful as moderating or mediating variables in the analysis of the impact.

The design of candidate analytics was guided both by the categories of implementation fidelity (e.g. adherence, exposure, quality of delivery, uptake) and by the pathways in the logic model. Below we describe how we used this guidance to design and try a wide variety of analytics.

1. *Adherence.* We were able to determine whether teachers were using the system to assign homework to their students, and whether they were appropriately using homework problems from their textbooks as well as “skill builder” problem sets. We could also see whether students were using the system to access and do homework.
2. *Exposure.* We could see how much homework was being assigned and how often it was assigned. We could see whether students were getting opportunities to use all ASSISTments features or just a subset of features. Another very interesting variable was the time of day when students were using ASSISTments: were they doing homework at home or in school as well?
3. *Quality of Delivery and Uptake.* For teachers, we could detect the “adaptive teaching” route of the logic model by seeing whether teachers were opening reports on their students’ homework (as opening the reports is a necessary precursor to adapting instruction on the basis of the reports). We also could detect student uptake and use of the system: how many minutes per week were students using the system? Was this consistent among students with a given teacher, or was it highly variable?

One important limitation of our plans is that teachers in the control schools are not using ASSISTments, and therefore we cannot get comparable data in the control schools. Because of this, the study is still using self-report, interview, and observational measures to get information about control schools and to understand the enacted contrast.

Data Sources and Analysis

In the 2012-2013 school year, 17 schools in the state of Maine were recruited as the first cohort of participants and 9 schools and all 7th grade teachers in these schools were randomly assigned to the treatment condition. Overall, 13 treatment teachers and over 800 7th grade students from their classes used ASSISTments to do their homework. Each teacher and student has his/her own login account. Thus, all actions made in the system can be tracked individually. As a student works online in ASSISTments, the system keeps a detailed log (aka “the click stream”) of his/her interaction with the tutor, including answers given, whether correct or incorrect, requests for hint messages, or other interface selections such as clicking on specific links to start an assignment or moving ahead to the next problem. Additionally, an offline version enables students to use ASSISTments even when they don’t have Internet access at home. Student work is recorded on their laptops and uploaded to the ASSISTments’ server when the laptop is connected to the Internet at school. The offline use data will be included in the reports when the teacher opens then online the next time. Teacher’s use of the system such as assigning homework, the type of the assignment, clicking a link to open a specific report, is also logged. All of the actions are time-stamped. To compute the candidate analytics, we collected ASSISTments system log data for the 13 teachers and their students from the period from February to April 2013. The log is fine-grained behavioral and outcome data as students interact with the system.

Measures of treatment fidelity were developed based on the log data to assess the extent to which teachers and students in the treatment condition followed the specified use model, as described above. Our approach to data analysis was essentially descriptive, using aggregated statistical metrics. At this stage of the efficacy trial, a goal was for the data analysis team to present a portrait of implementation to the development team, and to ask: *is this quality of implementation you were expecting to see and would be happy to have tested in with new cohort of student next year? If not, are there actions you can take that might bring implementation up to your desired levels before the next school year starts?*

Findings

A first useful analytic was how often teachers made assignments with ASSISTments. We found that across 3 months, teachers were assigning approximately 1-2 homework assignments per week in ASSISTments (Figure 3). This was lower than the 3 assignments per week that were originally expected. The team also looked at homework completion rates, which were around 75% and average minutes spent doing homework that was round 15 minutes. Both of these values were approximately as expected. Overall, the team felt the rate of homework assignment was a little low, but acceptable given the minutes spent doing homework and the completion rates.

We then looked at the type of assignment (See Figure 4). This revealed that teachers were assigning standard textbook homework problems about half the time and mastery “skill builders” about a quarter of the time. About 25% of the assignments were not from textbook. This was viewed as very promising, as it countered an earlier fear that teachers were sticking with traditional homework items and not using the potentially more powerful mastery problem sets in ASSISTments. We also learned from this analysis how useful analytic trend information can be. The teacher professional development logic for ASSISTments assumed that teachers would start with the more familiar “textbook homework” and gradually feel comfortable to include less familiar “skill builder” problem sets. The trend towards more “skill builder” usage suggested that this was indeed occurring.

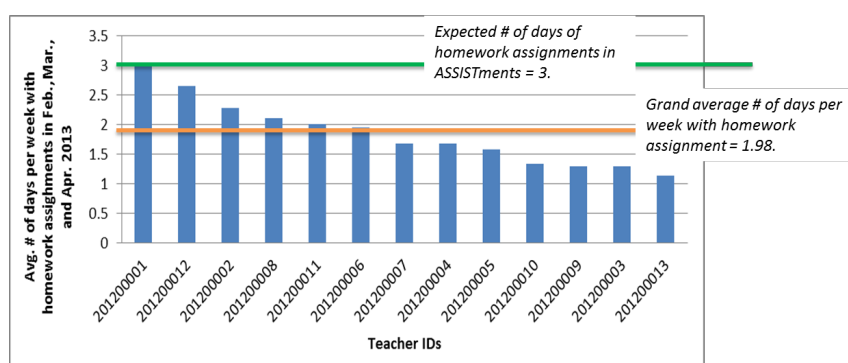


Figure 3. Teachers’ frequency of homework assignments in ASSISTments is lower than what was originally expected

School	Teacher ID	# of Textbook Problem Set assigned			Textbook Problem Set %	# of ASSISTments SkillBuilder assigned			SkillBuilder %	Total # of assignments
		Feb	Mar	Apr		Feb	Mar	Apr		
Clark Junior High School	201200006	12	18	6	69.23%	3	13		30.77%	52
Clark Junior High School	201200009	9	6		34.09%	22	3	4	65.91%	44
Clinton Elementary School	201200007	1	2	13	33.33%	12	6	14	66.67%	48
Clinton Elementary School	201200001	26	42	38	100.00%					106
Clinton Middle School	201200010	4	5	2	78.57%			3	21.43%	14
Clinton Middle School	201200004	4	4	3	40.74%	4	4	8	59.26%	27
District Middle School	201200012	15	16	10	59.42%	13	8	7	40.58%	69
Valley Middle School	201200003	10	13	2	100.00%					25
Clark Community School	201200011	5	6	1	75.00%		4		25.00%	16
Clark Community School	201200002	11	12	7	87.50%	1	3		12.50%	34
Clark Community School	201200013	3	4	1	57.14%	2	2	2	42.86%	14
William S. Cohen School	201200005	7	5	9	100.00%					21
William S. Cohen School	201200008	11	11	4	66.67%	10	3		33.33%	39
Grand Total					71.88%				28.13%	544

Figure 4. The type of assignments that teachers have made in ASSISTments

The graph in Figure 5 shows what percentages of problems were solved (on average) in each hour of the day. It is obvious that most use happens during the school time from 8am to 2pm, with some usage in after school hours from 3pm to 9pm. This was unexpected as the system was intended for homework analysis. Through complementary methods (such as teacher interviews), we are seeking to determine why most usage was during school hours.

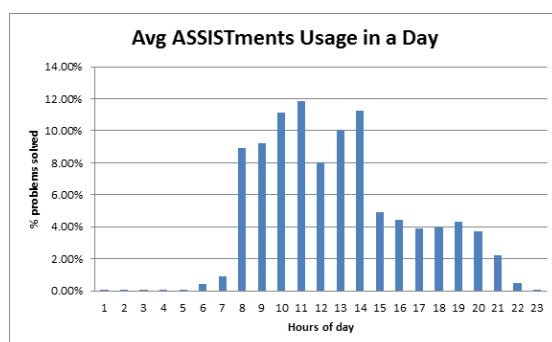


Figure 5. Distribution of student's usage of ASSISTments during a day

A key "uptake" analytic was whether teachers were opening ASSISTments reports, as this is a necessary prelude to adaptive teaching. Here, variation was profound and surprising. The key ASSISTments trainer was very surprised at the particular teachers who were not opening reports; apparently these teachers gave the impression that they were implementing adaptive teaching. We also saw variation within schools; different teachers in the same school were not using the reports equally often. This was confirmed by the field observations that are currently being conducted in schools. These data points led to concrete plans (such as a targeted discussion in an upcoming webinar, or a class visit) to follow up with the teachers who were not yet using the reports often, so as to activate the adaptive teaching pathway in the logic model for all classrooms in the treatment condition.

Conclusion

The quality of efficacy studies can be improved if the expected contrast between conditions can be actively maintained. This is typically difficult in large school-based studies, because of the difficulties associated with self-report (inaccurate), interviews (time consuming and unreliable), and observations (expensive). We explored the utility of creating analytics based on the automatically collected data in the ASSISTments system to address implementation fidelity.

We were able to map each of the three main pathways of the logic model to at least one analytic measure. The adaptive teaching pathway could be examined by looking at whether teachers open reports. The homework completion pathway could be examined by looking at homework completion rates. And the pathway of direct student impacts from having more support while doing homework could be examined by looking at the frequency of homework assignments, the types of problems included in the assignments, and the minutes students were spending doing assignments online. On a whole, this is an extraordinary amount of useful implementation data that we were able to gather and analyze at very modest cost and with high objectivity (especially compared to the inaccuracies of self-report and interview measures).

Further, we were able to design analytics corresponding to four of the five categories. We could look at adherence and exposure by whether teachers were using ASSISTments, how much homework was being assigned, and how many minutes of student usage. We could look at quality of implementation and uptake by seeing which types of problems teachers were assigning and whether they were opening reports. Trend data was particularly useful in showing that quality of implementation was improving over time, which was expected. The major limitation was that we did not have access to comparable data from control conditions, and thus the other more typical ways of collecting implementation data are still necessary.

ASSISTments team members who were coaching with teachers were able to target particular teachers and particular behaviors for their further coaching. Their surprise at which teachers needed coaching indicated the value of combining their own impressions with more objective analytic data. The ASSISTments team also learned from the time-of-day data that students were doing homework not just at night, but also during the school day and in the afternoon. Further, schools could be encouraged to set up school library computers for students to do homework during the day, which might further increase minutes spent on the system and completion rates.

In the future, we will be able to examine these analytics as mediating variables in our models of outcomes in the RCT. In the past, analytics have been predictive of outcome data. If the study finds that ASSISTments is efficacious, this could be very useful for making recommendations to schools and teachers for further implementation—we may learn that certain characteristics of usage best predict outcomes (such as number of minutes used) and this could guide schools in their further implementations. Considerable work remains to be done to more thoroughly validate particular analytic approaches; for example, it would be useful to compare interview or observational data to system-based measures. It could be, for example, that some teachers are assigning a mix of homework both within and outside ASSISTments, which could lead to different interpretations of how to intervene to increase implementation fidelity. Yet even at the descriptive level addressed here, the analytic data was perceived as very useful for working on the quality of implementation of the treatment prior to the year in which outcomes would be measured.

Overall, our recommendation is that evaluators who are planning RCTs to measure the efficacy of technology-based interventions consider how analytics could be used to measure implementation fidelity against the program logic model and across all categories of fidelity. The low cost, timeliness, and objectivity of analytics make it a valuable new tool—which can supplement traditional interview, observational, and self-report measures—and can lead to better control of the expected contrast between conditions. This, in turn, can improve the quality of an efficacy trial.

References

- Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York: John Wiley and Sons, Inc.
- Brown, M. (2011). *Learning analytics: The coming third wave*. EDUCAUSE Learning Initiative. Retrieved Nov. 2, 2013 from <http://net.educause.edu/ir/library/pdf/ELIB1101.pdf>.
- Cordray, D. S. (2008). *Fidelity of intervention implementation*. Paper presented at the IES 2008 Research Conference.
- Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-aged children. *Journal of Educational and Psychological Consultation*, 11, 193-221.
- Dron, J. and Anderson, T. (2009). On the design of collective applications, *Proceedings of the 2009 International Conference on Computational Science and Engineering*, Volume 04, pp.368-374.
- 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, 327-350.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W.B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18, 237-256.
- EDUCAUSE (2013a). *The NMC Horizon Report: 2013 Higher Education Edition*. Retrieved Nov. 5, 2013 from <http://net.educause.edu/ir/library/pdf/HR2013.pdf>.
- EDUCAUSE (2013b). *The NMC Horizon Report: 2013 K-12 Edition*. Retrieved Nov. 5, 2013 from <http://www.nmc.org/pdf/2013-horizon-report-k12.pdf>
- Fagan, A.A., Hanson, K., Hawkins, J.D., & Arthur, M.W. (2008). Bridging science to practice: Achieving prevention program implementation fidelity in the community youth development study. *American Journal of Community Psychology*, 41, 235-249.
- Feng, M. & Heffernan, N.T. (2007). Towards Live Informing and Automatic Analyzing of Student Learning: Reporting in the Assistment System. *Journal of Interactive Learning Research*. 18 (2), pp. 207-230. Chesapeake, VA: AACE.

- Feng, M., Heffernan, N. T., & Koedinger, K. R. (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 19(3), 243-266.
- Mendicino, M., Razzaq, L., & Heffernan, N. T. (2009). Comparison of traditional homework with computer supported homework: Improving learning from homework using intelligent tutoring systems. *Journal of Research on Technology in Education (JRTE)*, 41(3), 331-359.
- Mowbray, C.T., Holter, M.C., Teague, G.B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24, 315-340.
- Nelson, M C.; Cordray, D S.; Hulleman, C S.; Darrow, C L.; Sommer, E C. (2010). *A Procedure for Assessing Fidelity of Implementation in Experiments Testing Educational Interventions*. Society for Research on Educational Effectiveness. 2010.
- Pardos, Z. A., Baker, R. S. J. d., San Pedro, M. O. C. Z., Gowda, S.M., Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*, Leuven, Belgium. 117-124.
- Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., et al. (2007). *Organizing instruction and study to improve student learning*. IES practice guide (NCER 2007-2004). Washington, D.C.: National Center for Education Research. Retrieved February 28, 2008, from <http://ies.ed.gov/ncee/wwc/pdf/practiceguides/20072004.pdf>
- San Pedro, M. O. Z., Baker, R.S.J.d., Gowda, S.M., & Heffernan, N.T. (2013). Towards an understanding of affect and knowledge from student interaction with an intelligent tutoring system. *Proceedings of the 16th International Conference on Artificial Intelligence and Education*, Memphis, TN.
- San Pedro, M. O. Z., Baker, R.S.J.d., Bowers, A.J., Heffernan, N.T. (2013b). Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. *Proceedings of the 6th International Conference on Educational Data Mining*, Memphis, TN. pp. 177-184
- Singh, R., Saleem, M., Pradhan, P., Heffernan, C., Heffernan, N., Razzaq, L., & Dailey, M. (2011). Improving K-12 homework with computers. In *Proceedings of the Artificial Intelligence in Education Conference*, Auckland, New Zealand.
- U.S. Department of Education (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. Washington, D.C., 2012.

Acknowledgments

This material is based upon work supported by the Institute of Educational Sciences (IES) of U.S. Department of Education under Grant Number R305A120125. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES.