Exploring Issues of Implementation, Equity, and Student Achievement With Educational Software in the DC Public Schools

June Ahn

New York University

Austin Beck

University of Maryland, College Park

John Rice

Michelle Foster

District of Columbia Public Schools

In this article, we present analyses from a researcher-practitioner partnership in the District of Columbia Public Schools, where we are exploring the impact of educational software on students' academic achievement. We analyze a unique data set that combines student-level information from the district with data of student usage of a mathematics game platform: First in Math (FIM). These data offer a window into long-standing issues in the educational technology literature around implementation, equity, and student achievement. We show that time spent in FIM was correlated with improved future performance on standardized math assessments for students in Grades 4–8. However, student time spent using FIM was highly related to factors such as race, gender, and prior achievement. Such observations from data are helpful for school districts and researchers to inform equitable implementation of new technologies and maximize benefits to learners.

Keywords: *computers and learning, equity, mathematics education, regression analyses, technology*

RECENT studies have documented the rich ways that computers, software, games, online curriculum, and other new technologies affect how people access education, learn from one another, and develop new skills (Clark, Tanner-Smith, & Killingsworth, 2015; Ito et al., 2013; Means, Bakia, & Murphy, 2014). However, researchers, policy makers, and educators consistently face a seemingly intractable conundrum. Despite our hopes that new technologies will revolutionize education systems, with promising findings in controlled studies, we often find little positive impact on students' formal learning outcomes when technologies are implemented more broadly in schools or communities (Cuban, 1986; Vigdor & Ladd, 2010; Warschauer, Cotten, & Ames, 2011). This history looms large, as K–12 schools in the United States experience a new wave of optimism that technology can be deployed to enhance classroom practice and improve student achievement.

The prior literature in educational technology highlights the complicated relationship between technology and learning outcomes. A technological tool alone does not cause learning to occur (Clark, 1994). Instead, learning outcomes arise out of a complicated web of sociotechnical factors, including ensuring consistent access to hardware and software (Wood & Howley, 2012), attending to how teachers conceive of and implement technology (Cuban, Kirkpatrick, & Peck, 2001; Zhao & Frank, 2003), and understanding whether and how learners use the technology (Means et al., 2014). In studies of any given technology—ranging from computers and video games to online curriculum and wikis—scholars have also shown that social factors such as class, race, gender, and school environment are highly related to whether students have access to a new technology, engage with it, and use it in ways related to improved learning outcomes (Hohlfeld, Ritzhaupt, Barron, & Kemker, 2008; Reich, Murnane, & Willett, 2012; Shin, Sutherland, Norris, & Soloway, 2012). Issues of implementation and equity remain vital factors today, as K–12 school districts reinvest heavily in computer hardware and digital curriculum.

In this article, we present initial analyses of an urban school district's implementation of an online mathematics game called First in Math (FIM). We situate our work in two recent developments in education research. First, we are part of a researcher-practitioner partnership (RPP) where scholars and practitioners (a) jointly develop research questions and analyses to help districts improve their programs and implementation and (b) simultaneously contribute to academic research (Penuel, Fishman, Cheng, & Sabelli, 2011).

Our partnership is based in the District of Columbia Public Schools (DCPS), which in 2012 began to implement FIM in K–8. This implementation was part of a broader initiative in DCPS that had two goals: (a) to provide teachers with high-quality content during the rollout of the Common Core State Standards for math and (b) to expand learning with digital software across the district and improve mathematics achievement for specific subgroups (e.g., low-performing students, African American students). Our goals in this research are to directly inform practice (Gutiérrez & Penuel, 2014), but also shed light on critical theoretical issues regarding what happens when new technologies are implemented at scale across an urban district, the complex patterns of adoption that occur, and their relationships to student achievement.

Second, prior studies of technology implementation were built on individual qualitative case studies or surveys with teachers and school staff to reveal a rich array of findings about the complexity and challenges associated with technology implementation within schools (Cuban et al., 2001; Hohlfeld et al., 2008; Warschauer et al., 2011; Zhao & Frank, 2003). Recently, however, education researchers have increasingly become attuned to the opportunities around digital data sources, culled from the rising use of software and online tools, to examine questions of learning (Bienkowski, Feng, & Means, 2012; Martin & Sherin, 2013; Siemens & Long, 2011). We suggest that combining digital data sources from software and student information from school district data systems affords new ways to examine issues of technology implementation, equity, and student achievement that can provide actionable insights for our district partners and contribute to the research base.

Here, we use student-level data from DCPS, which contain demographic information (e.g., race, gender, grade level), along with formal measures of student achievement in standardized mathematics exams for >47,000 K–12 students. We then link this information with usage data that log the hours spent in FIM for each student in the district. By combining data on FIM usage with administrative and school-level information, our study illuminates how FIM was adopted across the city and how this adoption relates to future academic performance for students across the district. We find that time spent in FIM during the school year was correlated with improved performance on standardized math assessments at the end of the year. However, which students spent time in FIM was highly related to demographic and prior achievement indicators.

These findings illuminate an additional nuance to discussions of educational technology and equity. Policy makers and administrators often adopt technology with equality in mind—giving all students equal access to technology—but often fail to account for equity, which includes a focus on developing strategies and processes to equalize the benefits that all learners (not only the already privileged) can achieve with a new technology. We document how data analyses of technology use and student achievement can illuminate potential inequities across a district system, provide indicators for what may be going right or what areas need further attention, and help educators develop strategies for implementation that result in equitable outcomes.

Background: Technology, Implementation, and Equity

This research is part of a growing number of studies that examine the relationship between technology integration and student achievement in K–12 schools. Current initiatives use terms such as *blended learning*, which describes the ability for teachers to use face-to-face practices and offload some of the learning tasks to digital curriculum, games, or other learning software (Staker & Horn, 2012). Recent studies suggest that integrating digital tools into classroom settings may be beneficial for student learning outcomes. For example, a 2010 meta-analysis on online and blended learning found that students in blended learning conditions generally performed better than students in fully online or fully face-to-face conditions (Means, Toyama, Murphy, Bakia, & Jones, 2010). The authors of the study warn that positive achievement outcomes are not caused by technology alone and that other factors—namely, increased time spent on learning activities—may be the direct links to student outcomes.

Most studies of blended learning situations tend to examine higher education contexts, but research in the K–12 arena is steadily increasing. The results are mixed when different uses of educational software and student learning outcomes are examined (Barrow, Markman, & Rouse, 2008; D. B. Clark et al., 2015; Figlio, Rush, & Yin, 2013; Hong, Tsai, Ho, Hwang, & Wu, 2013; Jia et al., 2013; Pane, Griffin, McCaffrey, & Karam, 2014; Shin et al., 2012; Smith & Suzuki, 2015; Wijekumar, Meyer, & Lei, 2012). Most studies examine the efficacy or average effects of blended environments, as compared with a face-to-face or fully online treatment, and find positive benefits on student achievement. However, one study of 10,000 students and their usage of several software products found no significant difference between blended learning and fully face-to-face learning during the first year of implementation. The researchers found mixed results (some positive and some negative) during the second year of the study (Campuzano, Dynarski, Agodini, & Rall, 2009).

Prior research suggests that introducing computing, software, and games into formal instruction is associated with positive gains in student achievement, but there are open questions about how to explain these relationships, for whom, and under what conditions. For example, Means et al. (2010) suggest that learning gains are explained, not by technology, but by the added time spent with instruction or with the learning tasks that technology may enable. Furthermore, certain types of technology-driven blended learning situations may benefit subgroups of students more

than others. For example, Shin et al. (2012) found that White male students benefited the most from playing math games on the GameBoy system. These various findings point to the need to better parse out the role of technology in learning outcomes and the complex ways that implementation affects this relationship.

At the most basic level, students need access to a new technology or curricular tool and must readily use the tool to expect any learning benefits. Thus, we focus our analyses on a commonly collected metric in digital learning platforms— the time that a learner spent in a given platform. Through our RPP with DCPS, we were able to obtain detailed log data from the vendor of FIM. These data included variables that described when students logged into the platform and how much time they spent playing the math games in FIM throughout the school year. We recognize some of the limitations related to using time-on-task measures in the Discussion section (Kovanovic, Gašević, Dawson, Joksimovic, & Baker, 2016). However, taking into account these considerations, we note that having detailed access to time spent in FIM at the student level allows us to delve deeper into how software use was distributed across this urban school district and to explore the correlation between use and achievement more directly than what was possible in past research.

Past studies on technology implementation and equity in schools also inform our conceptual framework and data analyses. Education scholars know that there are many complex organizational and social processes that have to be aligned for technology to be used in effective ways (Fishman, Marx, Blumenfeld, Krajcik, & Soloway, 2004). School districts must make particular decisions about what technologies to purchase, and they must provide organizational support to maintain access to these resources. For effective adoption to take place within a given school, school leaders and teachers need to create environments such as peer advice networks, technical support staff, and organizational schedules conducive to experimenting with new technology (Cuban et al., 2001; Penuel et al., 2010; Warschauer et al., 2011; Zhao & Frank, 2003). Teachers need to be able to learn how to use technologies but, more important, learn how to effectively integrate tools to create better learning environments (Ertmer, 2005; Mishra & Koehler, 2006). Finally, personal factors—such as students' motivations, goals, prior learning histories, and life contexts—influence how they take up new technologies and for what means, such as play, socializing, and studying (Ahn et al., 2014; Azevedo, 2011; Ito et al., 2010; Ito et al., 2013; Polman, 2006). Thus, many factors may come into play in the interpretation of a simple metric such as time spent in the FIM platform.

In the following study, our RPP team focused on two research questions that examine issues of implementation, equity, and student achievement. First, we were interested to document and track which students were spending time in the FIM platform throughout the school year. Past research

in technology implementation suggests a stark inequality in terms of which students have access to and take up new technology (Warschauer & Matuchniak, 2010). DCPS structured its technology initiative to give hardware and software access to all schools in the district. However, there may still be inequitable uptake of FIM among students in different schools that is important to examine.

> *Research Question 1:* Were there differences in which students used FIM throughout the school year by indicators of gender, race, prior achievement, and school environment?

Second, an important question from the district perspective was to understand whether using FIM was correlated with improved student achievement at the end-of-year standardized assessments. We examined this question directly by exploring whether students' time spent in the software was related to student achievement measures.

> *Research Question 2:* Was students' time spent in FIM correlated with improved performance in standardized math assessments?

Method

DCPS is a large urban school district that, in the 2015–2016 school year, operated 113 schools serving P–12. The district enrolls >49,000 students. Approximately 88% of students identify with an ethnic minority group; 10% are English language learners; and 16% are in special education programs. Finally, >75% of students in the district qualify for free and reduced-price meals, which is a common indicator of students who fall under the poverty line. DCPS reported that in 2014 approximately 51% of its students scored "proficient" or "advanced" on the standardized math exam ("DCPS at a Glance," n.d.). However, the district, much like other urban school systems in the United States, sees stark gaps in educational achievement across indicators such as race and gender. For example, in 2014, the district reported that >90% of White students (female and male) in the district scored proficient in the math exam. Comparatively, only 45% and 37% of African American girls and boys respectively scored proficient on this assessment ("Achievement Gap, Race and Equity," n.d.).

In 2012, DCPS began to implement a mathematics program in Grades K–8 called FIM: a web-based series of mathematics games where students practice mathematics computation skills such as addition, subtraction, multiplication, and division, as well as concepts such as fractions, decimals, and integers. The activities focus on skill-building practice with the various mathematics concepts. It is important to note that the game tasks are largely akin to drilling math computation. For example, students may be presented with different multiplication problems and, in a timed

exercise, asked to fill in the answers to those problems, which are scored on correctness. Students can earn points and badges as they practice more. The program attempts to motivate students to engage through gamification elements, such as rankings and leaderboards (tracking a school's progress versus other schools), opportunities to win "stickers" that mark achievement, and chances to unlock bonus games.

It is also important to note the limitations of FIM's design. From a critical perspective, the program is not designed to teach deeper conceptual thinking or mathematical practice; instead, it is a program that presents drills and rote practice for students (Gee, 2005). From a learning perspective, the program can largely be viewed as a supplemental tool where students can develop skills through repeated practice, and we corroborated with our partners in DCPS that teachers largely used the program as extra practice activities for learners. DCPS offered the program to all schools, using an opt-in model of implementation rather than a top-down mandate.

Data Sources

Working with DCPS staff, we received student-level administrative information (which did not include personally identifiable information) and data on students' time spent in FIM throughout the 2012–2013 school year. We received data several times from DCPS, reflecting different time points between the 2011–2012 and 2012–2013 school years. Variables included in the administrative data sets were a student's grade level, gender, ethnicity, race, special education status, English language learner status, and schools attended. We also received student scores on the state standardized assessment called the DC Comprehensive Assessment System (DC-CAS), which students take at the end of each year between Grades 3 and 8 and once in grade 10. Starting in the 2014–2015 school year, DCPS adopted the PARCC assessment (Partnership for Assessment of Readiness for College and Careers), used by numerous states in alignment with the Common Core State Standards, and our future analyses will examine these achievement measures.

In partnership with FIM, we also received student-level data about usage of the online platform FIM. This data set included variables that allowed us to match unique IDs in the district's data with unique IDs in the FIM system. In addition, we received variables that allowed us to match students to schools. The log data of usage included details about which modules of FIM a student had completed at a given point in time, what stickers or achievements he or she had earned, and, critically for this analysis, how much time a student had spent in the system at different points in the school year. In this analysis, we focused on calculating the total time that students spent using FIM as an indicator for how much different students engaged with the program.

Data Analysis

To gain a holistic understanding of how FIM was used across the district, we first computed descriptive statistics of student characteristics in the district and general patterns of FIM usage. We then conducted regression analyses to delve deeper into specific patterns of the data. These descriptive analyses helped us to understand how FIM use was distributed across the system from a high level, and they informed more focused methodological decisions, such as narrowing the sample of students in our subsequent regression analyses.

Through this process, we decided to delimit the sample in regression analyses to all students in Grades 4–8 during the 2012–2013 school year, the first year that the DCPS implemented FIM. The population is limited to students in these grades because DCPS focused its implementation of FIM in Grades K–8, and consequently, we observed in the data that use of FIM sharply declined when a student reached the ninth grade. In addition, students do not start taking the DC-CAS exams until the third grade, thus eliminating any analyses of student achievement prior to that grade level. In our regression models (detailed below), we included a lagged variable of prior achievement, as well as aggregate measures of prior achievement at the school level as one indicator of school environment. Thus, fourth graders were the first students in the sample who had a measure of prior achievement (their third-grade test scores).

We also note that the distribution of time spent in FIM was highly skewed. Students in DCPS spent an average of about 4 hr using FIM over the entire school year. However, there was wide variance among students. Where many students never used the system (0 hr), some students logged as much as 121 hr in FIM over the school year. In general, we found that the most engaged students spent between 4 and 12 hr in the FIM games. We excluded any extreme outliers in our data set—namely, students who spent >50 hr in FIM. In this sample, students spent an average of 3.7 hr in FIM in the school year. Students in the 75th percentile spent about 4 hr, and those in the 90th percentile spent approximately 19 hr using the program.

Finally, an important detail is in the interpretation of DC-CAS performance. The DC-CAS raw scores are not scaled across grade levels; thus, one cannot validly compare a student's third-grade raw score on the exam with his or her raw score in fourth grade. To address this limitation, we standardized students' test scores (calculating their z scores) within a given year. Thus, we interpret changes in z score as a change in a student's relative standing as compared with one's peers in a given year, not as a raw score on the exam. In our regression models, we include only students with complete data ($N = 9,204$; Table 1) as compared with the 13,581 students in the entire population of Grades 4–8. However, in comparing both samples, we confirmed that the smaller sample with complete data closely reflected the general population of students in Grades 4–8.

TABLE 1

Summary Statistics for All Students in Grades 4–8 With Complete Data (N = 9,204)

Variable	*M (SD)* or %
FIM use, hr	3.698 (7.077)[a]
Female	48.9
ELL status	10.9
SPED status	19.4
Race/ethnicity	
Asian	2.0
American Indian	1.2
Black	73.4
Hispanic	10.5
Multiracial	2.6
Pacific Islander	0.3
White	10.1

Note. Data exclude 37 students who were extreme outliers (FIM use >50 hr). FIM = First in Math; ELL = English language learner; SPED = special education.
[a]Minimum = 0, maximum = 49.31

Regression Models

In our regression analyses, we first examine the relationship between student and school characteristics and students' time spent using FIM (Research Question 1). We include a variety of indicators that were available from the DCPS data to explore whether there were inequitable patterns of use across the district. For example, past studies (Shin et al., 2012) suggest that different software or games may be more readily taken up by boys; thus, we wanted to examine any gender patterns related to FIM use. We also wanted to explore any inequitable patterns of use by language status, special education status, and race. Finally, a common occurrence with the introduction of new technologies is that already privileged groups—for example, students who are already higher achieving academically—may more readily adopt and thus benefit from a new intervention (Toyama, 2015). As such, we examine whether students who have higher prior achievement or were enrolled in school environments where their peers also had higher achievement were more likely to use FIM. The following independent variables were included in our models:

Female: whether a student was female.
English language learner status: whether a student was designated as an English language learner.
Special education status: whether a student was enrolled in a special education program.
Race: the racial group of which the student was identified in the system. White students are the reference group in the regression models.
Elementary school: an indicator for whether the student was in an elementary school (K–5) or middle school (6–8).

Prior achievement: We included 2011–2012 DC-CAS performance (*z* score) as an indicator for prior achievement.

Prior literature suggests that within-school dynamics, such as leadership, teacher networks, and organizational routines, mediate how technology is used and, subsequently, how much time a student could spend with a piece of software (Cuban et al., 2001; Zhao & Frank, 2003). These measures of school culture were not collected by DCPS or readily available. However, we constructed an aggregate characteristic of school environment by taking a weighted average of all students' prior achievement as one indicator of a student's school environment.

School prior achievement: a weighted average of students' prior achievement in a school. We use this as one indicator of school environment (e.g., generally higher- versus lower-achieving peers).

The second research question examines whether time spent in FIM was correlated with future academic performance on the 2012–2013 DC-CAS. In this regression model, we include the same independent variables and a measure of time spent in FIM:

FIM usage: how many hours a student had spent in the system over the entire year (2012–2013) prior to taking the end-of-year DC-CAS exam.

Findings

Which Students Are Spending Time in FIM?

We present our analyses of students who use FIM in Table 2. The findings suggest a complex picture of which students are spending time with the FIM program. Elementary schools appear to be using FIM more often than middle schools (a finding corroborated anecdotally with our DCPS partners). Students in elementary schools spent approximately 2 more hr practicing skills in FIM versus those in middle schools. English language learners spent an average of 0.6 hr (36 min) more in FIM than their peers, while students designated in special education spent 0.6 hr less using the program. It appears that female students spent less time in FIM than their male peers (about 40 min less over the year).

Differences in racial groups were also significant. The findings suggest that students of color in DCPS are spending more time in FIM than their White peers. Asian and Black students spent >3 more hr in FIM over the course of the year; Hispanic students spent just under 3 more hr in FIM; and Pacific Islander students spent >5 hr more throughout the year. In addition, students who had higher prior academic achievement spent substantially more time in FIM in the

TABLE 2

Regression of Time Spent in FIM on Prior Achievement, Demographics, and School Prior Achievement

Variable	B	SE	p[a]
Prior academic achievement[b]	1.626	0.095	**.00**
Female	−0.662	0.144	**.00**
ELL	0.606	0.287	**.04**
SPED	−0.615	0.207	**.01**
Race/ethnicity			
Asian	3.79	0.567	**.00**
American Indian	1.23	0.695	.08
Black	3.23	0.280	**.00**
Hispanic	2.91	0.362	**.00**
Multiracial	1.83	0.499	**.00**
Pacific Islander	5.86	1.273	**.00**
Elementary school	2.18	0.144	**.00**
School prior achievement	−0.405	0.168	**.01**

Note. $N = 9,204$. $R^2 = 0.074$. Dependent variable: FIM use (hr). Race reference group: White. FIM = First in Math; ELL = English language learner; SPED = special education.
[a]Bold indicates significance ($p < .05$).
[b]2011–2012 DC Comprehensive Assessment System (z score).

school year. Our results suggest that a student who scored 1 *SD* higher in the prior year's DC-CAS exam spent 1.6 more hr using FIM. A final interesting finding was that students who were in schools that had higher prior academic performance in the DC-CAS (school prior achievement in Table 2) spent less time (approximately 25 min less) using FIM over the subsequent school year.

Taken together, these results offer hints at positive aspects of DCPS implementation of FIM and pose questions about aspects to improve. If we assume that FIM use is a positive activity (e.g., correlates to improved academic achievement, a question that we examine below), it is heartening to observe that students of color are exhibiting more use than their White peers. This finding is particularly salient in the DCPS context, as the district is focused on improving the academic performance of minority students, specifically Black students. In addition, we note that the negative coefficient for school prior achievement is a potential positive in our context. This finding suggests that lower-achieving schools in DCPS are making efforts to allow students more access and time with FIM, which could produce more equitable achievement over time—provided that this time is indeed positive for students' academic achievement. We note that the real-world differences in time are small here. Students in lower-achieving schools spent about 25 min more over the whole year (perhaps negligible in the real world, although statistically significant).

Our analysis of DCPS and FIM data also uncovered new questions for our partnership team. For example, why did female students spend less time in the program versus their male peers? Most significant, students who were already higher achieving were apt to spend more time in the program. This finding offers some questions about equity. If students who are already doing well use FIM more and then benefit subsequently in the next year, would these patterns of use lead to more inequitable student achievement gaps? Such findings also compel deeper questions about why certain students may choose to spend time with a technology while their peers reject them. For example, we note that FIM was designed for students to practice computation skills—or what the most skeptical of observers might label a *drill-and-kill* activity. It appears higher-achieving students were willing to spend more time with this activity. If there were valid reasons for low-achieving students to reject using FIM or for teachers serving these students to avoid adopting the technology, then future studies are needed to understand this phenomenon, rather than bluntly assuming that such students are unmotivated or that teachers are not implementing with fidelity.

In addition, the stakes may be lower in this case of FIM implementation because of the nature of this activity. Offering students some level of drill and practice may be beneficial; thus, the district could focus on making sure that lower-achieving students had some minimal level of time with the program. Although drill and practice are lower-level activities—ones that students should perhaps not be spending a majority of their time on—achieving some level of minimal but effective use for these activities is very feasible in this context. For example, if teachers of lower-performing students ensured that students spent only 15 min per week practicing drills in FIM, this would result in approximately 10 hr of time spent over the entire year: a minimal time spent on drill and practice, with more time devoted to higher-order mathematics learning, but perhaps still effective in helping improve student achievement.

Did Time Spent With FIM Correlate With Future Student Achievement?

The discussion of time spent with FIM and equitable implementation hinges on whether there is evidence that using the program was linked to improved student achievement. We present the results of regression analyses that explore whether FIM use was significantly associated with improved student performance on the end-of-year DC-CAS math assessment (Table 3). We find that an hour of FIM use was correlated with 0.007 *SD* of higher standing in the end-of-year DC-CAS, controlling for other variables. To explore the robustness of our findings, we also ran regression models with school-level fixed effects to account for the hierarchical grouping of students in schools and had similar results for all coefficients.[1] We note that students of color in DCPS scored substantially lower in the DC-CAS math assessment than their White peers, except for Asian

TABLE 3

Regression of 2012–2013 DC-CAS z Score on FIM Use, Prior Achievement, Demographics, and School Prior Achievement

| | *B* (*SE*) | |
Variable	Model 1	Model 2
FIM use, hr	**0.007** (0.001)**	**0.008** (0.001)**
Prior academic achievement[a]	**0.646** (0.008)**	**0.646** (0.008)**
Female	**0.028** (0.012)**	**0.029** (0.012)**
ELL status	–0.010 (0.025)	0.010 (0.024)
SPED status	**−0.241** (0.018)**	**−0.236** (0.018)**
Race/ethnicity		
Asian	–0.023 (0.048)	–0.019 (0.048)
American Indian	**−0.269** (0.059)**	**−0.266** (0.059)**
Black	**−0.275** (0.024)**	**−0.276** (0.024)**
Hispanic	**−0.138** (0.031)**	**−0.139** (0.031)**
Multiracial	**−0.114** (0.042)**	**−0.113** (0.042)**
Pacific Islander	**−0.216** (0.108)***	**−0.217** (0.108)**
Elementary school	0.021 (0.012)	0.022 (0.012)
School prior achievement	**0.176** (0.014)**	**1.94** (0.014)**
School Prior Achievement × FIM Use		**−0.008** (0.002)**

Note. $N = 9,204$. $R^2 = 0.662$. Dependent variable: 2012–2013 DC-CAS math assessment (*z* score). Race reference group: White. Bold indicates significance ($p < .05$). DC-CAS = DC Comprehensive Assessment System; FIM = First in Math; ELL = English language learner; SPED = special education.
[a]2011–2012 DC-CAS (*z* score).
*$p < .05$. **$p < .01$.

students, who did not differ. We also explored models with interaction terms for race and FIM use and found no significant findings. Spending time in FIM had similar benefits for all students by race.

If one were to extrapolate from this linear relationship, then using FIM for 10 hr during the whole school year (e.g., approximately 15 min per week for 40 weeks) would be related to an increase of about 0.07 *SD* in a student's standing relative to one's peers. Spending approximately 20 hr over the year would be related to about 0.14-*SD* increase in student standing. In Figure 1, we present a scatterplot of students' 2012–2013 DC-CAS math scores (*z* scores on the *y*-axis) and their time spent in FIM (*x*-axis). We see a small but steady positive relationship between hours spent practicing in FIM and performance on the DC-CAS. These findings suggest that a moderate level of use—say 10–20 hr for the whole school year which encompassed 90% of all students who used FIM—may be related to improved performance on the end-of-year assessment.

We also explored the relationship between time spent in FIM and academic performance for students in higher-performing school environments (as measured by the average academic achievement standing of the school). In Table 3, we present a regression model that includes an interaction term for School Prior Achievement × FIM Use. We find a negative relationship for the interaction term and DC-CAS performance on the end-of-year exam. This finding suggests that students who are enrolled in higher-achieving school environments (where the average achievement of students is
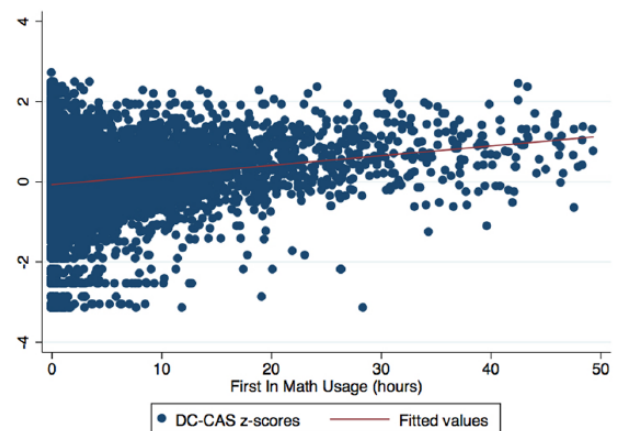

FIGURE 1. *Scatterplot of 2012–2013 DC Comprehensive Assessment System (DC-CAS) z scores (y-axis) and hours of use of First in Math (x-axis).*

higher) may not benefit as much from spending time with FIM, in terms of correlation with DC-CAS scores. One potential implication for this finding is that FIM may serve as a practice tool and that some level of practice may be related to improved performance on the standardized math assessment. However, students in schools with higher academic achievement environments may benefit more from other resources (different teachers, pedagogy, peers, activities, etc.), rendering the practice benefits afforded by FIM as not as beneficial.

Discussion

This study offers several implications for RPPs and research about implementation of educational software at scale in public education systems. We demonstrate the potential to combine student-level data from a school district, with usage data from digital platforms (e.g., FIM), to delve deeper into issues of technology implementation, student use, and relationships with learning outcomes of interest to educators and district leaders. Our analyses have been valuable for our RPP team, compelling new insights and further questioning that informs our future work together. This iterative inquiry-based process is core to RPPs (Penuel et al., 2011), and the availability of new data sources helps to facilitate this work. For example, our findings about students' time spent in FIM help to validate district decision making. We show that efforts of DCPS to provide computing, Internet, and technical infrastructure, as well as to make software such as FIM available to all schools, may have some positive benefits. Students of color show more time spent with FIM, and it appears that schools that serve lower-achieving students are providing students more access to the program (e.g., students in lower-achieving school environments are spending more time in FIM). Using FIM during the school year is also significantly correlated with higher performance on the DC-CAS at the end of the year. Thus, some amount of usage seems to be worth the district's efforts.

However, our analyses also suggest areas that require further inquiry for DCPS. For example, female students are spending less time in FIM, and students who are already higher achieving spend substantially more time in the program. These patterns compel further questions about those students who are not using FIM. Our data cannot tell us why these patterns of nonuse are present, but prior research provides some frameworks for potential avenues of future work in the district. Perhaps school leadership and organizational routines in different schools are shaping how FIM is presented and given to students for use (or nonuse). It is possible that the activities in FIM (drill and practice) are enjoyable for students who already have higher achievement but less so for students struggling to understand the math content and skills presented to them. Teachers could be conceptualizing their classroom practice in ways that may align or misalign with their understanding of how to use FIM with students. Our analyses are a first step in identifying potential inequities in implementation and spurring further questions for our RPP to understand these patterns in the data. We argue that this iterative inquiry—and a focus on equitable implementation patterns—would benefit school districts across the United States that are also implementing new digital platforms.

The findings suggest that spending time using FIM is correlated with improved performance on the DC-CAS at the end of the school year, but they also provide some nuance to how useful FIM may be and for which students. We know that FIM is limited in design, as largely a drill-and-practice application, and students spent relatively very little time over the course of the school year using it. However, even a small amount of drill and practice, such as 20 hr total over the entire school year, could potentially relate to higher performance on math assessments in the district (0.14-*SD* improvement). Our findings add to other studies showing that some computer-aided instructional programs could be related to improved academic achievement in math classrooms. For example, Barrow and colleagues (2008) found that a computer-aided math instruction platform—I Can Learn—was related to 0.17 *SD* of higher performance for middle and high school students.

Finally, we observed an interesting finding where spending time using FIM was more positive for students in lower-performing school environments. Taken together, these findings suggest some potential strategies for our RPP. It may be very reasonable for the district to frame FIM as a beneficial tool for students to practice basic skills, a little at a time and as a supplement, and focus the rest of its mathematics instruction on more ambitious learning activities. Past research in educational technology suggests that underserved and lower-performing schools are more likely to engage students in drill-and-practice forms of technology, with all of the negative connotations associated with that finding. However, our analyses suggest that a limited amount of drill and practice in FIM, for students in lower-performing school environments, appears to be beneficial for promoting student achievement on standardized assessments.

However, the limitations of our data also compel future questions for our RPP and general research on technology implementation in schools. First, time on task in a software environment is a potentially problematic measure for interpretation (Kovanovic et al., 2016). Our data cannot tell us how students spent time logged into the FIM platform—that is, whether they were progressing, distracted, logged in but not doing any tasks, and so on. We checked our time-on-task data with other indicators in the FIM log data to check for robustness. For example, we examined the relationship between the stickers/badges that players earned in FIM (for completing tasks) and their time spent in the program, and we found a largely linear relationship. This finding suggested to us that time also was a good indicator for progress (earning stickers/badges) in our FIM data set, and it gave us confidence to move forward with this measure in our models. However, in other districts or settings, this pattern is not a given, as time can be used in many different ways. From a theoretical perspective, we also cannot make claims about how students used their time in FIM beyond rote practice tasks. Did they speak with their peers, do activities together, or ask questions of their teachers? Our indicators for time provide analytics and patterns of interest to observe, but they are limited in delving into microlearning processes that are important.

Second, as with many field studies, we cannot control for the myriad of other programs that students are exposed to in the school district that are implemented simultaneously with FIM. For example, it may be possible that students with zero time on task in FIM or little time spent in the program are benefiting from other similar programs. Without data on those programs, it is difficult to fully pinpoint the impact of FIM on achievement. However, we previously noted that we examined models with school fixed effects and found similar results for time on task. In those models, we control for any programs adopted at the school level, and it would be likely that students within the same school would be exposed to similar sets of programs. That we find similar results for time on task in these models suggests some robustness of our findings.

Finally, we note that correlation does not equal causation. Our models take advantage of time, using strong controls such as prior achievement and modeling time spent in FIM to predict future academic performance. Although our correlational analyses provide compelling evidence that time spent in FIM was related to improved academic achievement in small but significant ways, we cannot make any causal claims that time spent in FIM caused better performance. It may be plausible that students who used FIM more also had better teachers, more motivation, academic supports, or other situations that explain the relationship between their time spent in FIM and academic achievement. From a theoretical perspective, we cannot rule out these alternative possibilities. However, from an analytics perspective, time spent in FIM may serve as a useful indicator that signals a need to look deeper into these students. Qualitative understanding about which students spend substantial time in the program and, relatedly, their peers who decide to be nonusers will be beneficial for the district to understand how to better support diverse learners over time. In this way, delving deeper into multiple data streams and analytics of digital trace data in educational software can play an integral role in school district decision making, educator practice, and improvement over time.

Note

1. We do not present these results given space constraints, but they can be provided by the first author upon request.

References

Achievement gap, race and equity. (n.d.). Retrieved from http:// dcps.dc.gov/publication/achievement-gap-race-and-equity

Ahn, J., Subramaniam, M., Bonsignore, E., Pellicone, A., Waugh, A., & Yip, J. C. (2014). "I want to be a game designer or scientist": Connected learning and developing identities with urban, African-American youth. In J. L. Polman, E. A. Kyza, D. K. O'Neill, I. Tabak, & W. R. Penuel (Eds.), *Proceedings of the Eleventh International Conference of the Learning Sciences (ICLS 2014)* (pp. 657–664). Boulder, CO: International Society of the Learning Sciences.

Azevedo, F. S. (2011). Lines of practice: A practice-centered theory of interest relationships. *Cognition and Instruction*, *29*(2), 147–184.

Barrow, L., Markman, L., & Rouse, C. E. (2008). *Technology's edge: The educational benefits of computer-aided instruction*. Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w14240

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

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts* (NCEE Report No. 2009-4041). Retrieved from the National Center for Education Evaluation and Regional Assistance website: http://eric. ed.gov/?id=ED504657

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2015). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*. Advance online publication. http://doi.org/10.3102/0034654315582065

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, *42*(2), 21–29.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York, NY: Teachers College Press.

Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal*, *38*(4), 813–834.

DCPS at a glance: Performance. (n.d.). Retrieved from http://dcps. dc.gov/node/1013812

Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development*, *53*(4), 25–39. http:// doi.org/10.1007/BF02504683

Figlio, D., Rush, & Yin, L. (2013). Is it live or is it Internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, *31*(4), 763–784. http://doi.org/10.1086/669930

Fishman, B., Marx, R. W., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *Journal of the Learning Sciences*, *13*(1), 43–76. http://doi.org/10.1207/s15327809jls1301_3

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, *2*(1), 5–16. http:// doi.org/10.2304/elea.2005.2.1.5

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher*, *43*(1), 19–23.

Hohlfeld, T. N., Ritzhaupt, A. D., Barron, A. E., & Kemker, K. (2008). Examining the digital divide in K–12 public schools: Four-year trends for supporting ICT literacy in Florida. *Computers & Education*, *51*(4), 1648–1663.

Hong, J.-C., Tsai, C.-M., Ho, Y.-J., Hwang, M.-Y., & Wu, C.-J. (2013). A comparative study of the learning effectiveness of a blended and embodied interactive video game for kindergarten students. *Interactive Learning Environments*, *21*(1), 39–53. http://doi.org/10.1080/10494820.2010.542760

Ito, M., Baumer, S., Bittanti, M., boyd, d., Cody, R., Herr-Stephenson, B., . . . Tripp, L. (2010). *Hanging out, messing*

around, and geeking out: Kids living and learning with new media. Cambridge, MA: MIT press.

Ito, M., Gutierrez, K., Livingstone, S., Penuel, W. R., Rhodes, J., Salen, K., . . . Watkins, S. C. (2013). *Connected learning: An agenda for research and design*. Retrieved from http://dmlhub. net/sites/default/files/ConnectedLearning_report.pdf

Jia, J., Chen, Y., Ding, Z., Bai, Y., Yang, B., Li, M., & Qi, J. (2013). Effects of an intelligent web-based English instruction system on students' academic performance. *Journal of Computer Assisted Learning*, *29*(6), 556–568. http://doi.org/10.1111/jcal.12016

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., & Baker, R. (2016). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110.

Martin, T., & Sherin, B. (2013). Learning analytics and computational techniques for detecting and evaluating patterns in learning: An introduction to the special issue. *Journal of the Learning Sciences*, *22*(4), 511–520. http://doi.org/10.1080/105 08406.2013.840466

Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when and how*. New York, NY: Routledge.

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: Office of Planning, Evaluation, and Policy Development.

Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor Algebra I at scale. *Educational Evaluation and Policy Analysis*, *36*(2), 127–144. http://doi. org/10.3102/0162373713507480

Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, *40*(7), 331–337. http://doi.org/10.3102/0013189X11421826

Penuel, W. R., Riel, M., Joshi, A., Pearlman, L., Kim, C. M., & Frank, K. A. (2010). The alignment of the informal and formal organizational supports for reform: Implications for improving teaching in schools. *Educational Administration Quarterly*, *46*(1), 57–95. http://doi.org/10.1177/1094670509353180

Polman, J. L. (2006). Mastery and appropriation as means to understand the interplay of history learning and identity trajectories. *Journal of the Learning Sciences*, *15*(2), 221–259.

Reich, J., Murnane, R., & Willett, J. (2012). The state of wiki usage in U.S. K–12 schools leveraging Web 2.0 data warehouses to assess quality and equity in online learning environments. *Educational Researcher*, *41*(1), 7–15.

Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology*, *43*(4), 540–560. http://doi.org/10.1111/j.1467-8535.2011.01197.x

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*, *46*(5), 30–32.

Smith, J. G., & Suzuki, S. (2015). Embedded blended learning within an algebra classroom: A multimedia capture experiment. *Journal of Computer Assisted Learning*, *31*(2), 133–147. http:// doi.org/10.1111/jcal.12083

Staker, H., & Horn, M. B. (2012). *Classifying K–12 blended learning*. Retrieved from the Innosight Institute website: http://www. innosightinstitute.org/innosight/wp-content/uploads/2012/05/ Classifying-K-12-blended-learning2.pdf

Toyama, K. (2015, May 19). Why technology will never fix education. *Chronicle of Higher Education*. Retrieved from http:// chronicle.com/article/Why-Technology-Will-Never-Fix/230185/

Vigdor, J. L., & Ladd, H. F. (2010). *Scaling the digital divide: Home computer technology and student achievement*. Retrieved from the National Bureau of Economic Research website: http:// www.nber.org/papers/w16078

Warschauer, M., Cotten, S. R., & Ames, M. G. (2011). One laptop per child Birmingham: Case study of a radical experiment. *International Journal of Learning and Media*, *3*(2), 61–76.

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://doi.org/10.3102/0091732X09349791

Wijekumar, K. K., Meyer, B. J., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Educational Technology Research and Development*, *60*(6), 987–1013.

Wood, L., & Howley, A. (2012). Dividing at an early age: The hidden digital divide in Ohio elementary schools. *Learning, Media and Technology*, *37*(1), 20–39.

Zhao, Y., & Frank, K. A. (2003). Factors affecting technology uses in schools: An ecological perspective. *American Educational Research Journal*, *40*(4), 807–840.

Authors

JUNE AHN is an associate professor at New York University, Steinhardt School of Culture, Education, and Human Development, 82 Washington Square East, New York, NY 10003; june.ahn@ nyu.edu. His research focuses on the design, implementation, and social/educational impact of learning technologies.

AUSTIN BECK is a doctoral student at the University of Maryland, College Park, 3119 Benjamin Building, College Park, MD 20742; austinbbeck@gmail.com. His research examines issues of educational technology, equity, and education policy.

JOHN RICE is the director of educational technology for secondary education, at the District of Columbia Public Schools, 1200 First St NE, Washington, DC 20002; john.rice@dc.gov.

MICHELLE FOSTER is the blended learning specialist for the District of Columbia Public Schools, 1200 First St NE, Washington, DC 20002; michelle.foster@dc.gov.