Empir Econ (2017) 52:1041–1063
DOI 10.1007/s00181-016-1189-y

The impact of computer-based tutorials on high school math proficiency

Ian K. McDonough[1] · Constant I. Tra[1]

Received: 7 January 2016 / Accepted: 10 August 2016 / Published online: 16 January 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract The benefits of mathematical-related skills are well documented in the economics and education literature. Even in spite of such evidence, proficiency levels among US high school students remain persistently low. This is especially true for the State of Nevada. As a result, the Clark County School District (CCSD) made available to students a computer-aided math tutorial prior to taking the High School Proficiency Exam (HSPE) in mathematics. As such, we utilize a novel dataset and explore the impact of computer-aided learning on mathematics proficiency rates for 10th and 11th graders in the CCSD. Our results provide some evidence of increased proficiency in mathematics related to tutorial participation. This is especially true for minority students. However, causal claims are limited due to the inability to rule out a zero lower bound on the estimated average treatment effects.

Keywords Academic achievement · Computer-aided education · Mathematical proficiency · Minimum biased estimator · Partial identification · Tutoring

JEL Classification H75 · I21

1 Introduction

The benefits of mathematical skills in school-age children are widely documented in the economics and education literature (Murnane et al. 1995; Grogger 1996; Heckman

✉ Ian K. McDonough
ian.mcdonough@unlv.edu

Constant I. Tra
constant.tra@unlv.edu

1 Department of Economics, University of Nevada, Las Vegas, 4505 S. Maryland Pkwy, Box 6005, Las Vegas, NV 89154, USA

المنارة للاستشارات

2006; Clements and Sarama 2011; Claessens and Engel 2013; Duncan and Magnuson 2013). However, mathematics proficiency levels among US high school students still remain low. In 2005, 2009, and 2013, only 23, 26, and 26% of US high school seniors were proficient in mathematics, respectively (Kena et al. 2015). On an international scale, US high school students (15-year-olds) currently rank 35 out of 64 countries in mathematics (OECD 2014).

At the state level, Nevada is one of the worst-performing states in the nation (Annie E. Casey Foundation 2016). In the largest school district within the state, the Clark County School District (CCSD), only 46% of 10th graders, on average, were deemed proficient in mathematics over the time frame spanning 2005–2009 (O'Brien et al. 2011).[1] As a way to help students prepare for the state-mandated High School Proficiency Exam (HSPE) in mathematics, the CCSD made available to students a computer-aided math tutorial prior to taking the exam in the 2006/2007 academic year. As such, this paper, using a novel dataset, explores the impact of computer-aided tutorials on proficiency levels in high school mathematics.

Much of the existing research focuses on computer-based tutorials as effective instructional methods in the classroom during school hours. The findings of these studies are mixed. Beal et al. (2007) use a controlled evaluation of an online tutoring system for high school math. The study finds that online-tutored students improved on the posttests but the impact was only limited to skills taught in the online system. Control group students showed no improvements. Barrow et al. (2009) use a randomized experiment to assess the impact of classroom software designed to help students with pre-algebra and algebra. The study finds that students assigned to computer-aided education score 0.17 standard deviations above students assigned to traditional classrooms on follow-up math tests.

While the aforementioned studies use locally administered tests to assess the impact of computer-aided education, other studies have investigated the impact on standardized tests scores. This is the case for Campuzano et al. (2009) where teachers in the same school were randomly assigned to use or not use a computer-assisted learning program. They find that students who were assigned to computer-aided classrooms did not show significant improvements in standardized test scores over the control group.

A few studies have used the meta-analysis approach to summarize the impacts of computer-aided learning in the classroom environment. Cheung and Slavin (2013) provide a meta-analysis of studies on the impact of technology in the classroom on mathematics achievements for K-12 students. They find that the use of technology in the classroom has a positive but modest effect compared to traditional methods. They also find that among the three types of educational technology applications, computer-aided instruction had the largest impact. Steenbergen-Hu and Cooper (2013) analyze 26 reports on intelligent tutoring systems spanning 1997 to 2010 and find that intelligent tutoring systems had *at worse* no negative effects on performance and even slightly positive. Additionally, when compared to face-to-face tutoring, the impact of computer-based tutoring had small to modest impacts. Kulik and Fletcher (2016)

[1] Using district data for the 2009/2010 school year, O'Brien et al. (2011) find that a 49.6% math proficiency rate for 10th graders in Clark County lags behind three comparable peer districts: Broward County (73.0%), Houston ISD (68.0%), and Miami-Dade County (73.0%).

المنارة للاستشارات

provide a recent meta-analysis of 50 studies on the impact of intelligent computer tutoring systems in the classroom on students' test scores. They find that the median effect among the 50 studies was to raise test scores by 0.66 standard deviations over traditional methods. However, the results were mixed among the studies with some finding very small effects. Their meta-analysis also revealed that the magnitude of the effects depended other factors such as whether the improvements were measured using standardized or locally administered tests, the nature of the control treatments, and the adequacy of program implementation.

Even fewer studies have investigated the effectiveness of computer-assisted tutorials that are used outside the classroom environment. This is an important aspect as such programs can be used as a complement for in-class learning in low-income areas where classroom technology may be lacking and the quality of the classroom instruction may fall short. A rare contribution on this question is Lai et al. (2015) who investigate the impact of computer-aided learning programs held out of regular classroom hours using a randomized experiment in Beijing, China. The authors find that the program improved students' standardized math scores by 0.15 standard deviations. The study also found that the program had the biggest impact on students with less-educated parents.

Given this prior literature, our motivation is twofold. First, there is very limited research exploring the impact that these tutorial systems have on math proficiency when viewed as complements outside of the classroom. Second, the prevailing results in the literature are quite mixed in terms of the benefits provided by computer-aided technology. As such, we contribute to the literature by focusing on 10th and 11th graders over the 2006/2007 school year in the CCSD and explore the impact of the *Succeed in Math!*® (SCIM) computer-aided tutorial on mathematics proficiency rates, as measured by performance on the HSPE in mathematics, when participation in the tutorial is *complementary* to conventional, in-class instruction. To mitigate the bias stemming from self-selection into the tutorial without access to valid exclusion restrictions given limitations of the data, we employ nonparametric bounding methods of Kreider et al. (2012) and a minimum biased (MB) treatment effect estimator proposed by Millimet and Tchernis (2013) to evaluate the *causal* effect of participation in the tutorial program.

Our results are noteworthy and provide some evidence that participation in computer-aided math tutorials, outside of school hours, increases the likelihood that the average high school student passes as proficient in mathematics. However, our ability to make definitive causal claims is limited since we cannot rule out a zero lower bound on the estimated average treatment effects. With that said, if we assume the average treatment effect (ATE) is given by the upper bound, then the impact on the average high school student in the CCSD would be to increase overall proficiency rates in our sample from 52 to 54%. For minorities the results are even larger. If every minority student in the sample were to participate in the math tutorial, and again assuming the upper bound to be the ATE, the mathematics proficiency rate for minority students in the sample would increase from 34 to 41%. If every male and female in the sample were to participate in the tutorial, the proficiency rates would increase from 55 to 58% and from 48 to 54% for males and females, respectively.

المنارة للاستشارات

The remainder of the paper is organized as follows. Section 2 briefly describes the SCIM tutorial. Section 3 describes the dataset used in the analysis. Section 4 presents the two estimators utilized. Section 5 discusses the results. Finally, Sect. 6 concludes.

2 The *S*ucceed in Math!® tutorial program

The SCIM software combines testing and tutoring services into one program that identifies a student's weaknesses in math and then creates a customized tutoring plan for that student. Furthermore, the software is designed to bring together teacher, student, and parent in the remediation process.[2]

The student begins by taking a short multiple-choice test that covers a gamut of topics as defined by the assessment standards set forth by the State of Nevada. The questions on the test can be categorized under four different content strands (numbers and operations, algebra and functions, geometry and measurement, and data analysis). Each strand can be further broken down into more specific concentrations. After the pretest, the program identifies those areas that the student needs to focus on, and a study plan is created from the program's library of content modules. If after the pretest a student is deemed "at grade level" in all content strands related to state proficiency standards, then no personalized tutorial program will be generated, and that student will be unable to proceed with the tutorial.

Conditional on having areas that need improvement, the student can then access the content modules and study at a pace to his or her own liking. A quiz is assigned after the completion of each module. If the student fails to pass a particular quiz, the program reassesses the student's abilities and introduces lower level content. As the student works through and masters this content, the program guides the individual back through the higher-level content until command of the topic is achieved. Upon completion of all modules, a posttest is assigned to measure improvement.[3]

3 Data

The dataset used in the analysis was provided by the CCSD. The dataset contains High School Proficiency Exam (HSPE) scores in mathematics for every student who took the exam in the school district over the 2006/2007 academic year. The HSPE score served a dual role at that time. First, it served as one of three exit-type subject exams that students must pass in order to successfully graduate high school in Nevada. If students do not pass as proficient on the HSPE in mathematics, they can still receive a certificate of attendance, but not a certified high school diploma. Second, as of the 2003/2004 academic year, the results for the HSPE in mathematics could be used to satisfy the upper-grade testing requirement set forth in No Child Left Behind (NCLB).[4]

[2] See http://www.succeedinmath.org/.

[3] Unfortunately, we do not have scores for either the tutorial pre- or posttest. Note, these pre- and posttest scores from the tutorial are *not* the HSPE scores (our outcome variable of interest).

[4] The institutional settings of NCLB in Nevada have evolved over the years. Nevada submitted an initial plan called "All Children Can Succeed" in June 2002 (Nevada Department of Education 2002). At the

المنارة للاستشارات

Our main outcome of interest, then, is a binary indicator for whether or not a student scores as proficient on the HSPE in mathematics. The exam was offered multiple times over the school year, with 96% of students in the sample taking the exam in March 2007. Since students can repeat the exam multiple times over the school year, we use the exam score that corresponds to the student's first attempt conditional on that first attempt not being classified as a "retest" exam score.[5] In addition to HSPE math scores, the dataset also contains information on whether each student participated in the SCIM tutorial program prior to taking the HSPE.

A limited set of student characteristics are also provided in the dataset. In particular, there is information on minority status (black or hispanic), gender, a binary indicator for English proficiency, a binary indicator for refugee status, and a binary indicator if the student was classified as gifted and talented. Unfortunately, there is no information provided on household-, parent-, and/or teacher-level characteristics. However, the refugee status indicator is picking up some information on the household since it is capturing whether or not a student is an English language learner. Meaning, a student can be classified as proficient in English yet speak some other language outside of school in the home. Nonetheless, the dataset is not ideal since we know some of these omitted factors matter in determining math (and reading) proficiency *and* are arguably correlated with participation in the tutorial program (Todd and Wolpin 2003, 2007). Our choice of estimators is guided by these data challenges.

Lastly, we drop students from the sample who dropped out of high school or were classified as special needs. Additionally, there were a group of students who selected into the tutorial, yet passed the tutorial pretest given at the onset of the program. Because these students passed the tutorial pretest, they were asked *not* to continue with the computer-aided tutorial since they were deemed already proficient. Further, because this group was neither strictly in the control nor treatment group, we drop these individuals from the final sample. Once we condition the dataset as described above, we are left with a sample of 15,360 10th grade students and 859 11th grade students for a total sample size of 16,219 students.

Table 1 provides detailed summary statistics for the full sample, as well as for the tutorial and non-tutorial participant subsamples. In terms of race, the tutorial participants appear very similar relative to non-tutorial participants. The one exception is for hispanics, where there are statistically fewer hispanics, proportionally, in the non-tutorial group relative to the tutorial group. Not surprisingly, then, there are also fewer,

Footnote 4 continued
onset, Nevada relied on the Terra Nova exam scores to define proficiency across various subjects, which was already in use for the statewide testing program. However, Nevada later implemented revised methods relying more on criterion-referenced exams to determine proficiency levels, hence the use of HSPE scores in mathematics (and other subjects) as a means of satisfying the upper-grade testing requirement in NCLB (Nevada Legislative Council Bureau 2005).

[5] We focus on the first attempt to keep the identification as clean as possible. In particular, we are concerned that with some students having taken the exam multiple times over a short period, outcomes can be impacted if learning has occurred with respect to material on the HSPE, the testing environment, etc. (see Matton et al. 2011). We further condition the exam scores on the first attempt *not* being a "retest" exam score. We chose not to include these "retest" students because it was unclear as to when they actually retested and what the environment was like in which they retested. However, we again conduct the analysis including these students, and the results are qualitatively similar.

المنارة للاستشارات

Table 1 Summary statistics

Variable	Full sample		Tutorial participants		Non-tutorial participants		Difference
	Mean	SD	Mean	SD	Mean	SD	
Proficient on state math exam (1 = proficient)	0.515	0.500	0.561	0.497	0.512	0.500	0.049*
White (1 = yes)	0.406	0.491	0.401	0.490	0.406	0.491	−0.005
Black (1 = yes)	0.139	0.346	0.157	0.364	0.138	0.344	0.019
Hispanic (1 = yes)	0.340	0.474	0.300	0.458	0.342	0.474	−0.042*
Gender (1 = female)	0.515	0.499	0.615	0.486	0.509	0.500	0.106*
Proficient in English (1 = yes)	0.770	0.500	0.786	0.410	0.769	0.421	0.017
English language learner (1 = yes)	0.091	0.288	0.069	0.254	0.092	0.289	−0.023†
Gifted and talented (1 = yes)	0.142	0.349	0.098	0.298	0.145	0.352	−0.047*
10th grader (1 = yes)	0.947	0.224	0.772	0.420	0.957	0.202	−0.185*

$N = 16,219$ for the full sample; of this, 908 are tutorial participants and 15,311 are non-participants
* $p < 0.01$
† $p < 0.05$
‡ $p < 0.10$

المنارة للاستشارات

in a proportional sense, English language learners among tutorial participants relative to non-tutorial participants. As well, a smaller proportion of females and gifted/talented individuals participated in the tutorial relative to non-participants. Lastly, there is a statistically significant difference in the proportion of 11th graders relative to 10th graders who participated in the tutorial. See Table 1 for more details.

4 Empirical methods

The objective of this analysis is to identify the causal effect of computer-aided math tutorials on student-level proficiency in mathematics. However, identifying the causal effect of tutorial participation is confounded when individuals self-select into the tutorial program on the basis of unobserved characteristics that are also correlated with math proficiency (i.e., a failure of the conditional independence assumption). A popular approach to dealing with this difficulty is to rely on an instrumental variable (IV) to isolate the exogenous variation in program participation. However, given the limitations of the dataset (as discussed above), there are no valid exclusion restrictions that can be relied on. Given these difficulties, we employ two different estimators to mitigate the impact of endogenous selection when valid exclusion restrictions are unavailable: the first is the nonparametric bounds approach of Kreider et al. (2012), and the second is the minimum biased approach proposed by Millimet and Tchernis (2013).[6]

4.1 Nonparametric bounds estimator

The nonparametric bounds approach of Kreider et al. (2012) requires weaker assumptions relative to other estimators employed to identify the average treatment effect (ATE) of program participation in the presence of endogenous selection. As such, the estimator yields sharp bounds on the ATE rather than point estimates. However, the bounds can still be informative given various assumptions governing the selection process.

Since our outcome of proficient versus non-proficient in mathematics is binary in nature, we follow closely to Kreider et al. (2012) and define the ATE as

$$\begin{aligned}\mathrm{ATE}\,(0,1|X\in\Omega) &= P(MP(1)=1|X\in\Omega)\\ &\quad - P(MP(0)=1|X\in\Omega),\end{aligned}\tag{1}$$

where MP is realized math proficiency, $MP(1)$ represents the math proficiency of a student if they participated in the math tutorial, $MP(0)$ represents the math proficiency of that same student had they not participated in the math tutorial, and $X\in\Omega$ are

[6] In applying the two methods, we use the Stata command *-tebounds-* for the nonparametric bounds approach of Kreider et al. (2012) and the Stata command *-bmte-* for the minimum biased estimator of Millimet and Tchernis (2013). See McCarthy et al. (2014) and McCarthy et al. (2015) for further details on the Stata commands.

المنارة للاستشارات

observable characteristics with values in the set Ω.[7] Intuitively, then, the ATE captures the average difference in math proficiency rates between those who participated in the math tutorial relative to those who did not.

As alluded to earlier, subjects were not randomly assigned into the tutorial (treatment) versus non-tutorial (control) groups. In absence of true randomization between the control and treatment groups, relying on the potential outcome $MP(1)$ as the counterfactual for all students who did not participate in the tutorial program, and relying on the potential outcome $MP(0)$ as the counterfactual for all students who did participate in the tutorial program, will lead to incorrect estimates of the ATE given the endogenous selection. To better understand this, and relying on the law of total probability, we can rewrite (1) as

$$\begin{aligned} P[MP(1) = 1] = P[MP(1) = 1|T^* = 1]P(T^* = 1) \\ + P[MP(1) = 1|T^* = 0]P(T^* = 0), \end{aligned} \tag{2}$$

where $T^* = 1$ for students who actually registered a positive amount of time in the math tutorial and $T^* = 0$ for students who spend zero amount of time in the tutorial. Since actual tutorial participation is observed, the sampling process identifies the probability of selecting into the treatment group, $P(T^* = 1)$, the probability of selecting into the control group, $P(T^* = 0)$, and the expected value of outcomes conditional of the outcome being observed, $P(MP = 1|T^* = 1)$. However, since we only observe one state of the world for each individual in the dataset, i.e., the sampling process does not shed light on $P[MP(1) = 1|T^* = 0]$, we cannot point identify $P[MP(1) = 1]$ solely on the basis of the sampling process.

4.1.1 Exogenous selection

As a benchmark, we start by considering the naïve case of assuming exogenous selection. The assumption of exogenous selection can be expressed by

$$P[MP(1) = 1|T^*] = P[MP(1) = 1],$$

where conditioning on observables in X is still implied. Further, given (2) we can write

$$\begin{aligned} P[MP(1) = 1] = P[MP = 1|T^* = 1] \\ P[MP(0) = 1] = P[MP = 1|T^* = 0]. \end{aligned}$$

As such, the ATE can be expressed as

$$\begin{aligned} \text{ATE} &= P[MP(1) = 1] - P[MP(0) = 1] \\ &= P[MP = 1|T^* = 1] - P[MP = 1|T^* = 0], \end{aligned} \tag{3}$$

[7] Like Kreider et al. (2012), we drop $X \in \Omega$ going forward for notational simplicity.

المنارة للاستشارات

which can be identified by the sampling process under the assumption that individuals were *correctly* classified into the control and treatment groups.[8]

4.1.2 No assumption on selection

We next consider the case of imposing no assumption on selection. Given the absence of any assumption governing the selection process, and under the assumption of no measurement error, the estimator yields the worst-case bounds (Manski 1995; Pepper 2000). With respect to the missing counterfactuals, the only thing we know is that they must lie within the unit interval since they represent latent probabilities. As such the lower and upper bounds on $P[MP(1) = 1]$ and $P[MP(0) = 1]$ can be identified by

$$P[MP = 1, T^* = 1] \leq P[MP(1) = 1] \leq P[MP = 1, T^* = 1] + P(T^* = 0)$$

and

$$P[MP = 1, T^* = 0] \leq P[MP(0) = 1] \leq P[MP = 1, T^* = 0] + P(T^* = 1),$$

respectively. The upper and lower bounds on the ATE can then be compactly represented by

$$B^U_{\text{ATE}} = B^U_{P[MP(1)=1]} - B^L_{P[MP(0)=1]} \tag{4}$$

and

$$B^L_{\text{ATE}} = B^L_{P[MP(1)=1]} - B^U_{P[MP(0)=1]}, \tag{5}$$

where B^U and B^L are the upper and lower bounds, respectively.

Note, the bounds under the worst-case scenario always include zero. As such, it is impossible to know whether or not participation in the tutorial program actually helped improve mean levels of proficiency in mathematics. In other words, we cannot sign the ATE under this scenario. However, as noted in Millimet and Roy (2015), some insight may be gleaned from the results since extreme values are excluded from the bounds.

4.1.3 Monotonicity assumptions

The bounds represented by (4) and (5) require no assumptions with respect to how students select into the tutorial program. As such, it is not possible to exclude zero from the bounds. In order to tighten the bounds on the ATE relative to the worst-case scenario, we make various monotonicity assumptions with respect to the relationship between proficiency in mathematics, participation in the tutorial program, and the underlying data. Specifically we consider the monotone treatment selection (MTS) assumption, the monotone instrumental variable (MIV) assumption, and, lastly, the

[8] Given the nature of the computerized tutorial, we precisely know who participated in the tutorial program since actual time spent in the program is digitally recorded. Thus, unlike Kreider et al. (2012) and Millimet and Roy (2015) who rely on self-reported participation, we do not allow for misclassification errors.

المنارة للاستشارات

monotone treatment response (MTR) assumption (Manski 1997; Manski and Pepper 2000).

Monotone treatment selection The MTS assumption formalizes the idea that students who participate in the computer-aided tutorial have, on average, better potential outcomes than those who do not participate in the tutorial. In addition to evidence of students positively selecting into tutoring programs (Gurun and Millimet 2008), this assumption conforms with intuition in that those students who stand to benefit the most from the program would be the ones actually selecting into the tutorial on the basis of these unobserved gains.

Following Kreider et al. (2012) and McCarthy et al. (2014), this assumption can be represented notationally as

$$P[MP(j) = 1|T^* = 1] \geq P[MP(j) = 1|T^* = 0] \tag{6}$$

for $j = 0, 1$ and where $MP(j) = 1$ represents better outcomes with respect to proficiency in mathematics. By imposing the MTS assumptions in (6), the bounds on the ATE are given by

$$B^U_{\text{ATE}} = \frac{P[MP = 1, T^* = 1]}{P(T^* = 1)} - \frac{P[MP = 1, T^* = 0]}{P(T^* = 0)} \tag{7}$$

and

$$B^L_{\text{ATE}} = P[MP = 1, T^* = 1] - \{P[MP = 1, T^* = 0] + P(T^* = 1)\}. \tag{8}$$

Monotone instrumental variable A further assumption to tighten the bounds exploits information through the use of a MIV. Unlike a traditional instrumental variable with mean independence, the MIV only requires a weakly monotone positive relationship between the instrumental variable, ν, and the potential outcomes (Manski and Pepper 2000). In other words, higher values of ν are, on average, associated with better potential outcomes. Notationally,

$$P[MP(j) = 1|\nu = \eta_1] \leq P[MP(j) = 1|\nu = \eta] \leq P[MP(j) = 1|\nu = \eta_2],$$

where $\eta_1 < \eta < \eta_2$ and $j = 0, 1$.

The MIV utilized in the analysis is a composite measure constructed from multiple binary indicators. To construct the MIV index, we follow Kling et al. (2007) and take an equal weighted average of individual z-scores where the z-scores are constructed using binary indicators for students that are proficient in English, white, and male. The choice of variables used to construct the index stems from the literature demonstrating a positive achievement gradient between mathematics and each of these measures (Hedges and Nowell 1995; Leahey and Guo 2001; Fryer and Levitt 2004, 2006, 2010; Hemphill and Vanneman 2011; Sohn 2012).

المنارة للاستشارات

Combining the MTS and MIV assumptions (MTS–MIV) yields

$$\sup_{\eta_1 \leq \eta} B^L(\eta_1) \leq P[MP(j) = 1 | \nu = \eta] \leq \inf_{\eta \leq \eta_2} B^U(\eta_2), \tag{9}$$

where B^U and B^L are again the upper and lower bounds, respectively, and $j = 0, 1$.[9] As outlined in McCarthy et al. (2014), to obtain final bounds on the ATE the sample is first split into k number of cells. The MTS bounds are then calculated for $P[MP(j) = 1]$, $j = 0, 1$, over the k cells resulting in B_k^U and B_k^L, $k = 1, \ldots, K$. Weighted averages of B^U and B^L are calculated across the k cells resulting in joint MTS–MIV bounds. Final bounds on the ATE are then derived using (4) and (5).

Monotone treatment response The last assumption imposed to tighten the bounds on the ATE is the MTR assumption, which requires math proficiency to be a monotonically increasing function in the treatment. In other words, the computer-aided tutorial cannot have a negative impact on the probability of being proficient in math.[10] While the argument that in class computer-aided learning may crowd out a student's time for traditional learning and hence lead to *worse* performance in math, the unique nature of the dataset requiring out of classroom time in the tutorial mitigates this concern. This is particularly true if one assumes that participating in the tutorial is at least as effective as any other tutoring/instructional methods available to students outside of the classroom. Essentially, given positive selection into the tutorial, combining the MTS and MTR (MTS–MTR) assumptions implies that the lower bound on the ATE becomes zero and the upper bound on the ATE is again expressed by (7).[11]

4.2 Minimum biased estimator

Unlike the bounding approach of Kreider et al. (2012), the minimum biased estimator (MB) of Millimet and Tchernis (2013) yields a point estimate. The MB estimator is an extension of the normalized inverse probability weighted (IPW) estimator of Hirano and Imbens (2001), which provides an unbiased estimate of the ATE when the conditional independence assumption holds. To minimize the bias stemming from the model being under-specified, the MB estimator trims the sample to only include those observations in some neighborhood around the biased minimizing propensity score, $P*$.

[9] See Proposition 1 in Manski and Pepper (2000).

[10] Steenbergen-Hu and Cooper (2013) review 26 different reports on intelligent tutoring systems spanning 1997 to 2010 and find that tutoring had at *worse* a nonnegative impact on achievement outcomes.

[11] As noted in McCarthy et al. (2015), combining the MTR assumption with the MTS–MIV assumptions is slightly more complicated since $P[MP(j) = 1]$, $j = 0, 1$ are bounded separately within each of the k cells. Since the MTR assumption requires that tutorial participation have no negative impact on the probability of being proficient in math, then it must be the case that B_k^L of $P[MP(1) = 1]$ must be strictly less than B_k^U of $P[MP(0) = 1]$ $\forall\, k = 1, \ldots, K$.

المنارة للاستشارات

As proposed in Millimet and Tchernis (2013), the MB estimator can be expressed as

$$\mathrm{ATE}\left[P^{*}\right]=\left[\sum_{i\in\Omega}\frac{MP_iT_i}{\hat{P}(X_i)}\Big/\sum_{i\in\Omega}\frac{T_i}{\hat{P}(X_i)}\right]-\left[\sum_{i\in\Omega}\frac{MP_i(1-T_i)}{1-\hat{P}(X_i)}\Big/\sum_{i\in\Omega}\frac{(1-T_i)}{1-\hat{P}(X_i)}\right], \tag{10}$$

where Ω is the subset of high school students with estimated propensity scores in some neighborhood around P^*, MP is still an indicator for math proficiency, T is again a treatment indicator, and X is the set of covariates summarized in Table 1, including gender interacted with race. Implementing the estimator empirically, we utilize a radius around P^*, θ, of 0.05 and 0.25. For example, $\theta = 0.05$ corresponds to 5% of both the control and treatment groups being contained in Ω. Thus, an increasing value of θ increases the efficiency of the estimator at the expense of increasing the bias. The opposite holds true for decreasing values of θ. Note, the interpretation of the estimated MB treatment effects needs to be done so carefully since the estimators are only utilizing observations around P^*. Although trimming the data minimizes the bias stemming from the unconfoundedness assumption failing, it does so at the expense of altering the interpretation of the parameter being estimated.

Lastly, Millimet and Tchernis (2013) further derive MB estimators for the ATE while relaxing the assumption of joint normality between the errors in the potential outcomes equations and in the treatment equation. As such, they present Edgeworth expansion (EE) versions of the estimators following Lee (1984). Given the simulation results presented in Millimet and Tchernis (2013), our preferred MB estimators are guided by whether or not the errors in the treatment equation are homoskedastic or heteroskedastic. Specifically, if the errors are homoskedastic, we report the EE versions of MB estimators for the ATE, and if the errors are heteroskedastic, we report the non-EE version of the MB estimator for the ATE.[12]

5 Results

In this section we present our results for the nonparametric bounds estimator and the MB estimator. We employ both estimators utilizing the full sample, as well as various subsamples. Specifically, we estimate each separately for male students, female students, minority (black or hispanic) students, and white students. Lastly, we discuss the results putting the estimates into more tangible measures.

5.1 Nonparametric bounds results

The results for the full sample and subsequent subsamples are given in Tables 2, 3, 4, 5, and 6. To reiterate, the treatment group is defined as those students who spent any

[12] Likelihood ratio tests for heteroskedasticity were conducted, and results are available upon request.

المنارة للاستشارات

Table 2 Sharp bounds on the ATE of math tutorial participation on math proficiency: full sample

Assumption	Estimated bounds	95% Confidence interval
Exogenous selection	[0.048, 0.048]	[0.007, 0.087]
No assumption on selection	[−0.508, 0.492]	[−0.517, 0.498]
MTS	[−0.508, 0.048]	[−0.517, 0.087]
MTS–MIV	[−0.508, 0.046]	[−0.517, 0.083]
MTS–MIV–MTR	[0.000, 0.046]	[0.000, 0.083]

Treatment is defined as spending any positive amount of time in the tutorial module; confidence intervals around the ATE are obtained following Imbens and Manski (2004) with 100 bootstrap repetitions; $N = 16,219$; see text for further details

Table 3 Sharp bounds on the ATE of math tutorial participation on math proficiency: males

Assumption	Estimated bounds	95% Confidence interval
Exogenous selection	[0.058, 0.058]	[0.014, 0.117]
No assumption on selection	[−0.545, 0.455]	[−0.554, 0.465]
MTS	[−0.545, 0.058]	[−0.554, 0.117]
MTS–MIV	[−0.545, 0.048]	[−0.554, 0.097]
MTS–MIV–MTR	[0.000, 0.048]	[0.000, 0.097]

Treatment is defined as spending any positive amount of time in the tutorial module; confidence intervals around the ATE are obtained following Imbens and Manski (2004) with 100 bootstrap repetitions; $N = 7860$; see text for further details

Table 4 Sharp bounds on the ATE of math tutorial participation on math proficiency: females

Assumption	Estimated bounds	95% Confidence interval
Exogenous selection	[0.055, 0.055]	[0.009, 0.093]
No assumption on selection	[−0.474, 0.526]	[−0.485, 0.537]
MTS	[−0.474, 0.055]	[−0.485, 0.093]
MTS–MIV	[−0.474, 0.055]	[−0.485, 0.093]
MTS–MIV–MTR	[0.000, 0.055]	[0.000, 0.093]

Treatment is defined as spending any positive amount of time in the tutorial module; confidence intervals around the ATE are obtained following Imbens and Manski (2004) with 100 bootstrap repetitions; $N = 8359$; see text for further details

positive amount of time in the tutorial program, and the control group is comprised of all other students who registered zero hours in the tutorial program.

Focusing on the results, three things become immediately apparent. First, in all cases, except for the subsample of white students, the point estimates under the assumption of exogenous selection are positive in magnitude and statistically significant. Under exogenous selection, minorities appear to get the most out of tutorial participation with a point estimate of 0.101. Or in other words, participation in the tutorial is associated with an average increase in the likelihood of being proficient

المنارة للاستشارات

Table 5 Sharp bounds on the ATE of math tutorial participation on math proficiency: minorities

Assumption	Estimated bounds	95% Confidence interval
Exogenous selection	[0.101, 0.101]	[0.054, 0.145]
No assumption on selection	[−0.351, 0.649]	[−0.364, 0.660]
MTS	[−0.351, 0.101]	[−0.364, 0.145]
MTS–MIV	[−0.349, 0.101]	[−0.359, 0.145]
MTS–MIV–MTR	[0.000, 0.101]	[0.000, 0.145]

Treatment is defined as spending any positive amount of time in the tutorial module; confidence intervals around the ATE are obtained following Imbens and Manski (2004) with 100 bootstrap repetitions; Minority students defined as either black or hispanic; $N = 7766$; $N_b = 2257$; $N_h = 5509$; see text for further details

Table 6 Sharp bounds on the ATE of math tutorial participation on math proficiency: whites

Assumption	Estimated bounds	95% Confidence interval
Exogenous selection	[−0.027, −0.027]	[−0.075, 0.018]
No assumption on selection	[−0.667, 0.333]	[−0.678, 0.347]
MTS	[−0.667, −0.027]	[−0.678, 0.018]
MTS–MIV	[−0.665, −0.027]	[−0.677, 0.018]
MTS–MIV–MTR	[0.000, −0.027]	[0.000, 0.018]

Treatment is defined as spending any positive amount of time in the tutorial module; confidence intervals around the ATE are obtained following Imbens and Manski (2004) with 100 bootstrap repetitions; $N_w = 6581$; see text for further details

on the state mathematics exam by 10.1 percentage points. The ATE under exogenous selection for females and males is 0.055 and 0.058, respectively.

Second, the monotonicity assumptions drastically tighten the upper bounds relative to the worst-case bounds (no assumption on selection). And third, in all cases we are unable to rule out negative effects from tutorial participation when relying simply on the joint MTS–MIV assumption. It is only when we invoke the MTR assumption that we are able to rule out negative effects.

Specifically focusing on the full sample, the assumption of positive selection reduces the bounds on the ATE from [−0.508, 0.492] to [−0.508, 0.048]. Utilizing our composite MIV further reduces the upper bound, albeit very negligibly, and as alluded to previously, we are unable to rule out negative effects. However, the joint MTS–MIV–MTR assumption further reduces the bounds to [0.000, 0.046] with a corresponding Imbens and Manski (2004) confidence interval of [0.000, 0.083].[13]

Looking at males and females individually, the same patterns emerge. For males, the assumption on selection decreases the worst-case bounds from [−0.545, 0.455] to [−0.545, 0.058]. The joint MTS–MIV bounds are further reduced to [−0.545, 0.048]. As with the full sample, we are only able to rule out negative impacts from tutorial participation by imposing the MTR assumption. The joint MTS–MIV–MTR bounds

[13] All Imbens and Manski (2004) confidence intervals reported are at the 95% level.

المنارة للاستشارات

are further reduced to [0.000, 0.048] with an Imbens and Manski (2004) confidence interval of [0.000, 0.097].

The results for females are very similar to the results for males. In particular, the assumption on selection reduces the worst-case bounds from [−0.474, 0.526] to [−0.474, 0.055]. However, unlike with males, imposing the joint MTS–MIV assumption does nothing to further reduce the bounds. The only way the bounds are tightened further is via the MTR assumption. Specifically, the joint MTS–MIV–MTR bounds are further reduced to [0.000, 0.055] with an Imbens and Manski (2004) confidence interval of [0.000, 0.093].

Turning to the results for minority and white students, we primarily focus on the results for minority students since the results for whites are statistically insignificant when imposing the joint MTS–MIV–MTR assumption. Specifically, the joint MTS–MIV–MTR bounds for white students are [0.000, −0.027] with an associated Imbens and Manski (2004) confidence interval of [0.000, 0.018]. For minority students, the pattern in the resulting bounds follows similarly to what has been described above for the full sample, males, and females. In particular, the assumption on selection significantly tightens the upper bound on the worst-case bounds. However, introducing an MIV does nothing to further reduce the upper bound, but immaterially tightens up the lower bound. Specifically, the joint MTS–MIV assumptions tighten the worst-case bounds from [−0.351, 0.649] to [−0.349, 0.101]. Further imposing the MTR assumption jointly with the MTS–MIV assumption further reduces the bounds to [0.000, 0.101] with a derived Imbens and Manski (2004) confidence interval of [0.000, 0.145].

In summary, using the partial identification approach of Kreider et al. (2012), we find evidence of an association between student participation in the computer-aided tutorial and passing as proficient the HSPE in mathematics. Although we are unable to rule out a zero average treatment effect, we are able to rule out very sizable effects. Additionally, we find no statistically relevant impact for white students. As a word of caution, these results hinge on the credibility of the underlying assumptions, especially the MTR assumption. Although the MTR assumption is admittedly a very strong one, there is ample evidence in the literature suggesting that participation in technology-related tutorial programs can, at *worse*, have *no* impact on student performance (Steenbergen-Hu and Cooper 2013).

5.2 Minimum biased results

The results for the full sample and subsequent subsamples are given in Table 7. The control and treatment groups are still defined as above. Unlike the results presented earlier, here we obtain a point estimate as opposed to upper and lower bounds. For comparison, we report the MB estimates along side the OLS and IPW estimates.[14] 95% empirical confidence intervals are also reported.

[14] The covariates utilized in each model are indicators for minority status, gender, English proficient, refugee status, gifted and talented, and for being in the tenth grade. Note, the exogenous selection results presented in the previous section correspond to OLS when regressing math proficiency on only a treatment indicator.

المنارة للاستشارات

Looking at the results for the ATE, the main points that become apparent are similar to those associated with estimated bounds. First, there is a positive and statistically significant impact of participation in the math tutorial, with the exception of the point estimates for males when $\theta = 0.05$ and for white students. Second, the point estimates for the ATE stemming from the MB estimator are lower than those from OLS, which conforms with the idea of positive selection into the tutorial on unobserved gains. Note, because of the bias/efficiency trade-off associated with θ (i.e., as θ gets larger, there is a trade-off for more bias for greater efficiency), our preferred results discussed here are those associated with $\theta = 0.05$. However, the results for $\theta = 0.25$ are given in Table 7. Lastly, all results need to be interpreted cautiously with respect to the biased minimizing propensity score, P^*.

Focusing on the full sample, the estimated ATE for participating in the tutorial is 0.078 ($P^* = 0.128$). Said another way, conditional on having a relatively mid to high rank in the estimated propensity score distribution of program engagement, participation in the tutorial is associated with an average increase of 7.8 percentage points in the likelihood that the average high school student passes the state high school mathematics exam (the median estimated propensity score is 0.047). This is less than the estimated ATE of 0.108 using both the OLS and IPW estimators.

Looking at males and females individually, the estimated ATE associated with program participation is 0.095 ($P^* = 0.065$) and 0.084 ($P^* = 0.139$) for males and females, respectively. As noted earlier, the estimated ATE for males is not statistically significant when $\theta = 0.05$. However, at $\theta = 0.25$ the estimated ATE for males is again 0.095, but is now marginally statistically significant at the 95% confidence level. The estimated impact of tutorial participation for the average female is statistically significant at $\theta = 0.05$. For males, then, conditional on having a relatively mid to high rank in the estimated propensity score distribution of tutorial participation, participating in the tutorial is associated with a 9.5 percentage point increase in the likelihood of passing the HSPE in mathematics (the median estimated propensity score is 0.035). For females, again conditional on having a relatively mid to high rank in the estimated propensity score distribution of tutorial participation, participating in the tutorial is associated with a 8.4 percentage point increase in the probability of being classified as proficient on the HSPE in mathematics (the median estimated propensity score for females is 0.059).

Focusing on the results for minority and white students, it becomes immediately clear that the estimated ATE for white students is not statistically significant at either $\theta = 0.05$ or $\theta = 0.25$. This result is not surprising given the results from the estimated bounds presented above. However, for minority students we do find a positive and statistically significant association between participating in the tutorial and passing the HSPE in mathematics. Specifically, conditional on having a relatively high rank in the estimated propensity score distribution of tutorial participation, participating in the tutorial is associated with a 9.4 percentage point increase in the probability of being classified as proficient on the HSPE in mathematics ($P^* = 0.193$; the median estimated propensity score for minorities is 0.037).

To recap, implementing the MB estimator of Millimet and Tchernis (2013), we find some evidence of a positive, causal relationship between student participation in the computer-aided tutorial and passing as proficient on the high school state exam

المنارة للاستشارات

Table 7 Impact of math tutorial participation on math proficiency: full sample and student subgroups

	Average treatment effect (ATE)				
	Full sample	Males	Females	Minorities	Whites
OLS	0.108	0.097	0.115	0.144	0.065
	[0.082, 0.137]	[0.057, 0.137]	[0.085, 0.149]	[0.104, 0.181]	[0.021, 0.103]
IPW	0.108	0.078	0.129	0.145	0.044
	[0.081, 0.139]	[0.036, 0.130]	[0.096, 0.169]	[0.104, 0.195]	[0.000, 0.008]
MB $\theta = 0.05$	0.078	0.095	0.084	0.094	-0.099
	[0.027, 0.149]	[−0.008, 0.166]	[0.020, 0.166]	[0.010, 0.198]	[−0.176, 0.109]
$\theta = 0.25$	0.141	0.095	0.084	0.175	0.023
	[0.043, 0.169]	[0.006, 0.156]	[0.027, 0.167]	[0.047, 0.240]	[−0.019, 0.109]
$P*$	0.128	0.065	0.139	0.193	0.020
	[0.043, 0.350]	[0.020, 0.562]	[0.114, 0.188]	[0.020, 0.711]	[0.020, 0.434]

Treatment is defined as spending any positive amount of time in the tutorial module; 95% empirical confidence intervals in brackets obtained using 250 bootstrap repetitions; OLS—ordinary least squares estimator; IPW—inverse probability weighted estimator; MB—minimum biased estimator; P*—biased minimizing propensity score. For the list of covariates used, see Table 1 in the text

المنارة للاستشارات

Table 8 Impact of math tutorial participation on math proficiency: bivariate probit sensitivity analysis

	Correlation of disturbances					
	$\rho = 0.00$	$\rho = 0.05$	$\rho = 0.10$	$\rho = 0.15$	$\rho = 0.20$	$\rho = 0.25$
Full sample	0.328*	0.223*	0.117†	0.011	−0.096†	−0.204*
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.046)
	[0.106]	[0.072]	[0.038]	[0.003]	[−0.032]	[−0.067]
Males	0.290*	0.181†	0.072	−0.039	−0.150†	−0.261*
	(0.076)	(0.076)	(0.076)	(0.075)	(0.075)	(0.074)
	[0.093]	[0.059]	[0.023]	[−0.013]	[−0.050]	[−0.087]
Females	0.353*	0.249*	0.146†	0.042	−0.063	−0.167*
	(0.061)	(0.061)	(0.060)	(0.060)	(0.060)	(0.059)
	[0.114]	[0.081]	[0.047]	[0.014]	[−0.020]	[−0.054]
Minorities	0.445*	0.338*	0.232*	0.125‡	0.018	−0.089
	(0.069)	(0.069)	(0.069)	(0.068)	(0.068)	(0.067)
	[0.153]	[0.116]	[0.079]	[0.042]	[0.006]	[−0.029]
Whites	0.187†	0.082	−0.024	−0.130‡	−0.238*	−0.347*
	(0.075)	(0.075)	(0.075)	(0.075)	(0.074)	(0.074)
	[0.057]	[0.026]	[−0.008]	[−0.042]	[−0.078]	[−0.116]

Degree of selection on unobservables defined by ρ; robust standard errors in parentheses; average partial effects in brackets; for a list of covariates used, see Table 1

* $p < 0.01$

† $p < 0.05$

‡ $p < 0.10$

in mathematics. To reiterate, one needs to be cautious when interpreting these results since the sample (subsamples) is (are) trimmed to only include students around the bias minimizing propensity score.

5.3 Sensitivity analysis

To assess the sensitivity of our results to selection on unobservables, we follow Altonji et al. (2005) and estimate a bivariate probit model varying the degree of selection on unobservables. In particular, we estimate

$$\begin{aligned} T &= \mathbb{I}(X'\beta + \upsilon > 0) \\ MP &= \mathbb{I}(X'\gamma + \alpha T + \varepsilon > 0), \end{aligned} \tag{11}$$

where T is again equal to one for students who participate in the tutorial, *MP* is again equal to one for students who test proficient in mathematics, X is the same set of covariates as earlier, and

$$\begin{bmatrix} \upsilon \\ \varepsilon \end{bmatrix} \sim \mathbb{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

المنارة للاستشارات

while constraining ρ to particular values. Specifically, we vary ρ from 0 to 0.25 in increments of 0.05. When $\rho = 0$ there is no selection on unobservables; however, each subsequent iteration through values of $\rho > 0$ allows for more positive selection into the treatment on the basis of unobservable factors. We report both the probit coefficients and average partial effects on mathematics proficiency probabilities in Table 8.

What becomes immediately apparent when looking at the results is that the impact of tutorial participation on the probability of passing as proficient in mathematics is very sensitive to selection on unobservables. Across the full sample and various subsamples, even just a moderate amount of selection on unobservables can erase away any positive impacts resulting from tutorial participation. For the full sample, when $\rho = 0$, the impact on the margin for tutorial participation is 0.106. However, when $\rho = 0.15$ the estimated impacts are no longer statistically significant at conventional levels. When ρ is increased to 0.20, the estimated coefficient turns negative with a value of -0.096 and is again statistically significant at conventional levels. A very similar pattern emerges when looking at the results for the various subsamples. With the exception of minorities, the average partial effect becomes negative and/or statistically insignificant at conventional levels when $\rho = 0.15$. For minorities, statistical significance is lost when $\rho = 0.20$. Again, detailed results are given in Table 8.

5.4 Discussion

By imposing restrictions on how students self-select into the math tutorial program and by trimming the sample to obtain minimum biased point estimates, the results presented here provide *some* evidence that participation in computerized math tutorials increases the likelihood of testing as proficient in high school math. However, the results need to be interpreted cautiously since we are unable to rule out a zero lower bound on the ATE estimates. Overall, the results lend some support for the limited evidence that finds a positive impact of computer-aided learning programs implemented *outside* of regular school hours (Lai et al. 2015).

In framing the results in more intuitive terms, we focus on the ATE and rely on the bounding estimator of Kreider et al. (2012) since these bounds, relative to the MB point estimates, provide more conservative estimates for the impact that tutorial participation has on the average high school student's proficiency in mathematics. In particular, using the estimated upper bound for the full sample under the joint MTS–MIV–MTR assumption, and assuming the impacts are causal, if every student participated in the tutorial, then the overall math proficiency rate in the sample could potentially increase from 52 to 54% (or an additional 343 students in the sample would, on average, pass the HSPE in math).

The impacts are even greater when simply looking at minority students. Specifically, if every minority student were to participate in the tutorial, and again assuming the impacts from tutorial participation are causal, the math proficiency rate for minorities in the sample could potentially increase from 34 to 41% (or an additional 490 minority students would, on average, pass the HSPE in math).

المنارة للاستشارات

The impacts for males and females individually fall between the impacts presented above for the full sample and minority subsample. Specifically, for males, participation in the tutorial has the potential of increasing the math proficiency rate for males in the sample from 55 to 58% (or an additional 161 male students would, on average, pass the HSPE in math). If every female were to participate in the tutorial, then proficiency rates for this subsample could potentially increase from 48 to 54% (or an additional 196 female students would, on average, pass the HSPE in math).

Referring back to the results for the full sample, and using an estimated cost of $5.00 per tutorial license paid by CCSD, we can provide a back-of-the-envelope calculation for the per additional proficient student cost of the tutorial program.[15] Relying on the upper bound on the ATE for the full sample, and again assuming that the impacts are in fact causal, the estimated cost to achieve an additional 343 proficient students in mathematics is approximately $236 per student. Recent figures suggest that high school graduates make 33% more in lifetime earnings relative to someone with no high school diploma (or roughly $9000 less per year) (Carnevale et al. 2011). As such, and assuming that not passing the HSPE is what keeps a non-proficient student from graduating high school, the cost of $236 per additional proficient student seems inconsequential compared to the potential benefits accrued over the life cycle. Again as a word of caution, these impacts are obtained using the upper bounds on the ATE stemming from participation in the math tutorial. So in essence, under the MTS–MIV–MTR assumption, the impacts can range anywhere from *zero* to the figures reported above.

As a final note, if we were to perform the same exercise using the MB point estimates, the potential impacts would only be larger. This, of course, is conditional on the MB results being generalizable across the entire propensity score distribution of tutorial participation. However, given the results from the sensitivity analysis, even just a modest amount of selection on unobservables can explain any positive impact stemming from tutorial participation.

6 Conclusion

In this paper we present results on the impact of computer-aided tutorials on math proficiency rates when the tutorial utilized is a *complement* to traditional classroom instruction in that the computer-aided tutorial is implemented outside of normal school hours. We use data on students from the Clark County School District in Nevada during the 2006/2007 school year and match up tutorial participation with corresponding test scores on the state proficiency exam. Our choice of estimators is guided by the inherent, and arguably severe, limitations of the dataset. Given the limited set of relevant covariates as it relates to achievement in mathematics, the conditional independence assumption surely fails. Identifying the impact of tutorial participation on proficiency in mathematics is further confounded by not having access to valid exclusion restric-

[15] The cost of $5.00 per license comes from conversations with individuals at CCSD who recall the program during the 2006/2007 school year. The numbers provided to us were in the range of $4.00 to $5.00 per license. We opt to use the more aggressive $5.00 per license to construct an upper bound on the cost per additional proficient student.

المنارة للاستشارات

tions. As such, we employ the partial identification methods of Kreider et al. (2012) and the minimum biased estimator of Millimet and Tchernis (2013).

Overall our results provide some evidence that tutorial participation increases math proficiency for the average high school student. This is particularly true given our MB estimates. However, we are unable to rule out a zero average treatment effect using the partial identification approach of Kreider et al. (2012). Further, any positive impact stemming from tutorial participation appears to be highly sensitive to selection on unobservables. Taken together, the results presented here provide some, although limited, support for the small literature on computer-aided learning implemented *outside* of traditional school hours providing a positive impact on achievement in mathematics (Lai et al. 2015). Given the benefits of mathematical proficiency documented in the economics and education literature, policy makers and school districts alike may find it beneficial, and possibly cost effective, to combat low proficiency rates in mathematics by implementing computer-aided tutorials as a *complement* to traditional classroom instruction.

Acknowledgements The authors have benefited from helpful suggestions from the co-editors, Wenhua Di and Daniel Millimet, and two anonymous referees, as well as participants at the *Intent versus Impact: Evaluating Individual- and Community-Based Programs* conference hosted by the Federal Reserve Bank of Dallas.

References

Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. J Polit Econ 113:151–184

Annie E. Casey Foundation (2016) KIDS COUNT data book, 2016: state trends in child well-being. The Annie E. Casey Foundation, Baltimore

Barrow L, Markman L, Rouse CE (2009) Technology's edge: the educational benefits of computer-aided instruction. Am Econ J Econ Policy 1:52–74

Beal CR, Walles R, Arroyo I, Woolf BP (2007) On-line tutoring for math achievement testing: a controlled evaluation. J Interact Online Learn 6:43–55

Campuzano L, Dynarski M, Agodini R, Rall K (2009) Effectiveness of reading and mathematics software products: findings from two student cohorts (NCEE 2009-4042). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education

Carnevale AJ, Rose SJ, Cheah B (2011) The college payoff: education, occupations, lifetime earnings. Technical report, Georgetown University Center on Education and the Workforce

Cheung ACK, Slavin RE (2013) The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: a meta-analysis. Educ Res Rev 9:88–113

Claessens A, Engel M (2013) How important is where you start? early mathematics knowledge and later school success. Teach Coll Rec 115:1–29

Clements DH, Sarama J (2011) Early childhood mathematics intervention. Science 333:968–970

Duncan GJ, Magnuson K (2013) Investing in preschool programs. J Econ Perspect 27:109–132

Fryer RG, Levitt SD (2004) Understanding the Black–White test score gap in the first two years of school. Rev Econ Stat 86(2):447–464

Fryer RG, Levitt SD (2006) The Black–White test score gap through third grade. Am Law Econ Rev 8(2):249–281

Fryer RG, Levitt SD (2010) An empirical analysis of the gender gap in mathematics. Am Econ J Appl Econ 2(2):210–240

Grogger J (1996) Does school quality explain the recent Black/White wage trend? J Labor Econ 14:231–253

Gurun A, Millimet DL (2008) Does private tutoring payoff? IZA working paper no. 3637

المنارة للاستشارات

Heckman JJ (2006) Skill formation and the economics of investing in disadvantaged children. Science 312:1900–1902
Hedges LV, Nowell A (1995) Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science 269:41–45
Hemphill FC, Vanneman A (2011) Achievement gaps: how Hispanic and White students in public schools perform in mathematics and reading on the national assessment of educational progress (NCEE 2011-459). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, Washington, DC
Hirano K, Imbens G (2001) Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Serv Outcomes Res Methodol 2:259–278
Imbens G, Manski C (2004) Confidence intervals for partially identified parameters. Econometrica 72:1845–1857
Kena G, Musu-Gillette L, Robinson J, Wang X, Rathbun A, Zhang J, Wilkinson-Flicker S, Barmer A, Dunlop Velez E (2015) The condition of education 2015. U.S. Department of Education Report (NCES 2015-144), National Center for Education Statistics, Washington, DC
Kling JR, Liebman JB, Katz LF (2007) Experimental analysis of neighborhood effects. Econometrica 75:83–119
Kreider B, Pepper JV, Gundersen C, Jolliffe D (2012) Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported. J Am Stat Assoc 107:958–975
Kulik JA, Fletcher JD (2016) Effectiveness of intelligent tutoring systems: a meta-analytic review. Rev Educ Res 86:42–78
Lai F, Luo R, Zhang S, Huang X, Rozelle S (2015) Does computer-assisted learning improve learning outcomes? evidence from a randomized experiment in migrant schools in Beijing. Econ Educ Rev 47:34–48
Leahey E, Guo G (2001) Gender differences in mathematical trajectories. Soc Forces 80(2):713–732
Lee LF (1984) Tests for the bivariate normal distribution in econometric models with selectivity. Econometrica 52:843–863
Manski C (1995) Identification problems in the social sciences. Harvard University Press, Cambridge
Manski C (1997) Monotone treatment response. Econometrica 65:1311–1334
Manski C, Pepper J (2000) Monotone instrumental variables: with an application to the returns to schooling. Econometrica 68:997–1010
Matton N, Vautier S, Raufaste E (2011) Test-specificity of the advantage of retaking cognitive ability tests. Int J Sel Assess 19:11–17
McCarthy I, Millimet DL, Tchernis R (2014) The bmte command: methods for the estimation of treatment effects when exclusion restrictions are unavailable. Stata J 14:670–683
McCarthy I, Millimet DL, Roy M (2015) Bounding treatment effects: stata command for the partial identification of the average treatment effect with endogenous and misreported treatment assignment. Stata J 15:411–436
Millimet DL, Roy M (2015) Partial identification of the long-run causal effect of food security on child health. Empir Econ 48:83–141
Millimet DL, Tchernis R (2013) Estimation of treatment effects without an exclusion restriction: with an application to the analysis of the school breakfast program. J Appl Econ 28:982–1017
Murnane RJ, Willett JB, Levy F (1995) The growing importance of cognitive skills in wage determination. Rev Econ Stat 77:251–266
Nevada Department of Education (2002) The Nevada plan: 'all children can succeed': consolidated plan for the implementation of the no child left behind act. State of Nevada Department of Education, Carson City
Nevada Legislative Counsel Bureau (2005) Report to the 73rd Session of the Nevada Legislature by the Legislative Committee on Education. Legislative Counsel Bulletin No. 05-15. Legislative Counsel Bureau, Carson City
O'Brien B, Brown S, Eno J, Pontoni K (2011a) Educational and operational efficiency study: analysis of student performance for the Clark County school district. Technical report, American Institute for Research
O'Brien B, Tucker N, Zuber T (2011b) Educational and operational efficiency study: comparative analysis of academic performance for the Clark County school district. Technical report, American Institute for Research

المنارة للاستشارات

OECD (2014) PISA 2012 results in focus: what 15-year-olds know and what they can do with what they know. OECD Publishing, Paris

Pepper J (2000) The intergenerational transmission of welfare receipt: a nonparametric bounds analysis. Rev Econ Stat 82:472–488

Sohn K (2012) A new insight into the gender gap in math. Bull Econ Res 64(1):135–155

Steenbergen-Hu S, Cooper H (2013) A meta-analysis of the effectiveness of intelligent tutoring systems on K-12 students' mathematical learning. J Educ Psychol 105:970–987

Todd PE, Wolpin KI (2003) On the specification and estimation of the production function for cognitive achievement. Econ J 113:3–33

Todd PE, Wolpin KI (2007) The production of cognitive achievement in children: home, school, and racial test score gaps. J Hum Cap 1:91–136

المنارة للاستشارات

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

www.manaraa.com