NEW TECHNOLOGY AND TEACHER PRODUCTIVITY

Eric S. Taylor[†]
Harvard University

January 2018

I study the effects of a labor-replacing computer technology on the productivity of classroom teachers. In a series of field-experiments, teachers were provided computer-aided instruction (CAI) software for use in their classrooms; CAI provides individualized tutoring and practice to students one-on-one with the computer acting as the teacher. In mathematics, CAI reduces by one-quarter the variance of teacher productivity, as measured by student test score gains. The reduction comes both from improvements for otherwise low-performing teachers, but also losses among high-performers. The change in productivity partly reflects changes in teachers' decisions about how to allocate class time and teachers' effort.

JEL No. I2, J2, M5, O33

[†] eric_taylor@harvard.edu, Gutman Library 469, 6 Appian Way, Cambridge, MA 02138, 617-496-1232. I thank Eric Bettinger, Marianne Bitler, Nick Bloom, Larry Cuban, Tom Dee, David Deming, Caroline Hoxby, Brian Jacob, Ed Lazear, Susanna Loeb, John Papay, Sean Reardon, Jonah Rockoff, Doug Staiger, and seminar participants at UC Berkeley, University of Chicago, Harvard University, UC Irvine, Stanford University, and University of Virginia for helpful discussions and comments. I also thank Lisa Barrow, Lisa Pithers, and Cecilia Rouse for sharing data from the ICL experiment, the Institute for Education Sciences for providing access to data from the other experiments, and the original research teams who carried out the experiments and collected the data. Financial support was provided by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090016 to Stanford University; and by the National Academy of Education/Spencer Dissertation Fellowship Program.

Computers in the workplace have, broadly speaking, improved labor productivity.[1] The productivity effects of computers arise, in part, because workers' jobs change: computers replace humans in performing some tasks, freeing workers' skills and time to shift to new or different tasks; and computers enhance human skills in other tasks, further encouraging reallocation of labor (Autor, Katz, and Krueger 1998; Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011). In this study I measure the effects of a labor-replacing computer technology on the productivity of classroom teachers. My focus on one occupation—and a setting where both workers and their job responsibilities remain fixed—provides an opportunity to examine the heterogeneity of effects on individual productivity.

Whether and how computers affect teacher productivity is immediately relevant to both ongoing education policy debates about teaching quality and the day-to-day management of a large workforce. K-12 schools employ one out of ten college-educated American workers as teachers,[2] and a consistent empirical literature documents substantial between-teacher variation in job performance.[3] In recent years, these differences in teacher productivity have become the center of political and managerial efforts to improve public schools. Little is known about what causes these differences, and most interventions have focused either on changing the stock of teacher skills—through selection or training—or on

[1] See for example Jorgenson, Ho, and Stiroh (2005), Oliner, Sichel, and Stiroh (2007), and Syverson (2011).

[2] Author's calculations from Current Population Survey 1990-2010.

[3] Much of the literature focuses on teacher contributions to academic skills, measured by test scores. In a typical result, students assigned to a teacher at the 75th percentile of the job performance distribution will score between 0.07-0.15 standard deviations higher on achievement tests than their peers assigned to the average teacher (Jackson, Rockoff, and Staiger 2014). Other work documents variation in teachers' effects on non-test-score outcomes (Jackson 2014), and teacher' observed classroom practices (Kane, McCaffrey, Miller, and Staiger 2013). Recent evidence suggests that variability in performance contributes to students' long-run social and economic success (Chetty, Friedman, and Rockoff 2014b).

changing teacher effort—through incentives and evaluation.[4] Computer technology is both a potential contributor to observed performance differences and a potential intervention to improve performance, but, to date, it has received little attention in the empirical literature on teachers and teaching.[5]

Two features of most classroom teaching jobs are important to predicting the effects of computers on individual productivity, and these features make heterogeneous effects more likely. First, the job of a teacher involves multiple tasks—lecturing, discipline, one-on-one tutoring, communicating with parents, grading, etc.—each requiring different skills to perform.[6] The productivity effects of a new computer which replaces (complements) one skill will depend on the distribution of that particular skill among the teachers. The effects of a labor-replacing technology will further depend on how the teacher's effort and time, newly freed-up by the computer, are reallocated across the tasks which remain the responsibility of the teacher herself. Second, teachers have substantial autonomy in deciding how to allocate their own time and effort, and the time and effort of their students, across different tasks. In other words, individual teachers make meaningful educational production decisions in their own classrooms. Differences in these choices likely explain some of the baseline variability in teacher productivity, even conditional on teacher skills. And, when a new labor-replacing computer becomes available, teachers themselves will partly decide how effort and time are reallocated. These two features are not unique to teaching, however,

[4] For examples from the literature on teacher selection see Staiger and Rockoff (2010), and Rothstein (2012). For training see Taylor and Tyler (2012). For incentives and evaluation see Barlevy and Neal (2012) and Rockoff, Staiger, Kane and Taylor (2012).

[5] There is some theoretical work on this topic. Acemoglu, Laibson, and List (2014) show how technology could permit productivity-enhancing specialization in teacher job design. Lakdawalla (2006) and Gilpin and Kaganovich (2011) consider how economy-wide technological change affects selection of people into and out of the teacher labor market by changing the relative skill demands in other sectors. Barrow, Markman, and Rouse (2008, 2009) discuss how technology could increase the quantity of instructional time.

[6] By "skills" I mean teachers' current capabilities whether innate, or acquired by training or experience, or both.

and so the analysis in this paper should have applicability in other occupations (see for example Atkin et al. 2017). The theoretical framework in Appendix B describes, in greater detail, the salient features of a teacher's job, the teacher's educational production problem generally, and the introduction of a new technology.[7]

In this paper I analyze data from a series of randomized field experiments in which teachers were provided computer-aided instruction (CAI) software for use in their classrooms. I first estimate the treatment effect on the variance of teacher productivity, as measured by contributions to student test score growth. I then examine whether the software affected individual teachers' productivity differentially, and examine the extent to which the software changed teachers' work effort and decisions about how to allocate time across job tasks.

Computer-aided instruction software effectively replaces teacher labor. It is designed to deliver personalized instruction and practice to students one-on-one, with each student working independently at her own computer and the computer taking the role of the teacher. Most current CAI programs adaptively select each new lesson or practice problem based on the individual student's current understanding as measured by previous practice problems and quizzes.[8] The experiments collectively tested 18 different CAI software products across reading in grades 1, 4, and 6; and for math in grade 6, pre-algebra, and algebra.

I report evidence that, among math teachers, the introduction of computer-aided instruction software reduces by approximately one-quarter the variation in

[7] I propose a version of the teacher's problem that (i) makes a clear distinction between the *tasks* that comprise the job of a classroom teacher, and a teacher's *skills* in each of those tasks; and (ii) explicitly considers the teacher's own decisions about education production in her classroom. The task-skills distinction is a useful and increasingly common feature in the literature on how technical change affects labor (Acemoglu and Autor 2011).

[8] A distinction is sometimes made between computer-*aided* and computer-*managed* instruction, with the latter reserved for software which includes the adaptive, individualized features. For simplicity and following prior usage in economics, I refer to this broader category as computer-aided instruction or CAI.

teacher productivity, as measured by student test scores. The standard deviation of teacher effects among treatment teachers was 0.22 student standard deviations, compared to 0.30 for control teachers. The reduction in variance is the result of improvements for otherwise low-producing teachers, but also losses in productivity among otherwise high-producing teachers. However, estimates for reading teacher productivity show no treatment effects.

The sign of the effect on variance is likely consistent with most reader's priors. If a computer skill replaces teacher skill in performing a given task, then the between-teacher variation in the productivity of that particular task should shrink. However, skill substitution in the given task is only the first-order effect. The total effect of some new technology on the variance of teacher productivity will depend on how individual teachers choose to reallocate time and effort across other tasks after giving some task(s) to the computer (see Appendix B for more discussion of this point and the next two paragraphs).

I also find evidence that the new software changes how teachers' carry out their job day-to-day. Data from classroom observations show a substantial reallocation of class time across tasks: treatment teachers increase by 35-38 percent the share of class time devoted to individual student work (often work using the CAI software), with offsetting reductions in the share of class time in whole-class lectures. This reallocation is consistent with teachers making a rational production decision: spending more of their class-time budget on individual student work and less on lectures because CAI increases the marginal rate of technical substitution of the former for the latter in producing student achievement. The reallocation is further motivated by a change in the relative effort costs. CAI reduces teacher effort on two margins. First, the teacher's role during individual student practice time shrinks to mostly monitoring instead of actively leading. Second, treatment math teachers reduce their total work hours, cutting time previously spent on planning and grading in particular.

Additionally, the reduction in effort costs, especially at the labor-leisure margin, is one explanation for why high-performing teachers might rationally choose to begin using CAI even though it reduces their student's achievement scores. Consistent with this explanation, as detailed below, the labor-leisure shift is largest among the relatively high-performing teachers. Willingness to trade student achievement for reduced own effort adds important nuance to the notion of teachers as motivated agents (Dixit 2002).

For most results in the paper, the argument for a causal interpretation relies only the random assignment study designs. This is the case for the reduction in the variance of teacher productivity, and the average changes in teacher effort and time allocation.[9] I use unconditional quantile regression methods to estimate the treatment effect heterogeneity. Some strong interpretations of quantile treatment effects require a rank invariance assumption. However, even if this assumption does not hold, the results still support important causal conclusions about the heterogeneity of effects, including the conclusion that productivity improved for some otherwise low-performing teachers but declined for some high-performers.

The analysis in this paper suggests new computer technology is an important contributor to differences in teacher productivity.[10] It also highlights interactions between teachers' skills and teachers' production decisions in determining observed performance.[11] Replacing teacher labor with machines, like

[9] Subtly, while the direction and magnitude of change in the variance of productivity are identified by random assignment alone, identifying the *level* of variance requires a further assumption, i.e., the standard identifying assumption about student sorting common throughout the teacher value-added literature. I discuss this issue later in the paper.

[10] Jackson and Makarin (2016) provide experimental evidence from another empirical example: providing lesson plans as a substitute for teacher effort and skill. As with CAI, the effects depend on prior teacher performance. Previously low-performing teachers improved, while there was little to no effect for high-performing teachers.

[11] Examination of teachers' production decisions by economists has been rare (Murnane and Phillips 1981; Brown and Saks 1987; and Betts and Shkolnik 1999 are exceptions).

the computer-aided instruction example I examine, can greatly benefit students in some classrooms, especially the classrooms of low performing teachers, while simultaneously making students in other classrooms worse off. This difference in outcomes arises partly because, given the option, some teachers choose to use a new technology, even if it reduces their students' achievement, because it also substantially reduces their workload.

1. Computers in schools and computer-aided instruction

Research evidence on whether computers improve schooling is mixed at best. Hundreds of studies take up the question—often reporting positive effects on student outcomes—but a minority of studies employ research designs appropriate for strong causal claims. That minority find mixed or null results (see reviews by Kirkpatrick and Cuban 1998; Cuban 2001; Murphy et al. 2001; Pearson et al. 2005). In the economics literature, several studies examine variation in schools' computer use induced by changes in subsidies (Angrist and Lavy 2002; Goolsbee and Guryan 2006; Machin, McNally, and Silva 2007; Leuven, Lindahl, Oosterbeek, and Webbink 2007; Barrera-Osorio and Linden 2009). In these studies, schools respond to the subsidies by increasing digital technology purchases, as expected, but with no consistent effects on student outcomes. In broad cross-sectional data, Fuchs and Woessmann (2004) find positive correlations between computers and student outcomes, but also demonstrate that those relationships are artifacts of omitted variables bias.[12]

[12] Evidence on the educational benefits of home computers is also mixed. Fuchs and Woessmann (2004), Vigdor and Ladd (2010), and Malamud and Pop-Eleches (2011) all find negative effects of home computers. In a recent field-experiment, Fairlie and Robinson (2013) find no effect of a computer at home on achievement, attendance, or discipline in school. By contrast, Fairlie (2005), Schmitt and Wadsworth (2006), Fairlie, Beltran, and Das (2010), and Fairlie and London (2012) all find positive effects.

Of course, “computers in schools” is a broad category of interventions. Computers can contribute to a range of tasks in schools: from administrative tasks, like scheduling classes or monitoring attendance, to the core tasks of instruction, like lecturing and homework. Today, software and digital products for use in schools is a nearly eight billion dollar industry (Education Week 2013). In this paper, I focus on one form of educational computer technology—computer-aided instruction software—which is designed to contribute directly to the instruction of students in classrooms.

1.A Description of computer-aided instruction

Computer-aided instruction (CAI) software is designed to replace traditional teacher labor by delivering personalized instruction and practice problems to students one-on-one, with each student working largely independently at her own computer. Most CAI programs adaptively select each new tutorial or practice problem based on the individual student’s current understanding as measured by past performance on problems and quizzes. If the student has yet to master a particular concept, the software teaches that concept again. Most products provide detailed reports on each student’s progress to teachers.

Figure 1 shows screen images from two different CAI products included in the data for this paper. As the top panel shows, from software for use in an algebra class, some CAI products largely replicate a chalkboard-like or textbook-like environment, though the product shown does actively respond in real-time with feedback and help as the student enters responses. The bottom panel, from a first grade reading lesson, shows one frame from a video teaching phonics for the letters l, i, and d. With its animated characters and energetic tone of voice, the latter is, perhaps, an example of the often cited notion that computers can provide a more “engaging” experience for students.

1.B Evidence on the student achievement effects of computer-aided instruction

While CAI was a new option for (most) teachers in this study, CAI is not a new technology. The psychologist B. F. Skinner proposed a “teaching machine” in the 1950s, and the development and research evaluation of computer-aided instruction dates back to at least the mid-1960s. Early experimental studies documented positive, often large, effects on student achievement (Suppes and Morningstar 1969; Jamison, Fletcher, Suppes, and Atkinson 1976).

In the past decade, results on CAI have been decidedly more mixed, again, especially if one focuses on studies with rigorous designs (see review in Dynarski et al. 2007). Many field-experiments testing several software programs find zero effects of CAI (or at least null results) on student test scores in reading and math classes at elementary and secondary school levels (for reading see Rouse and Krueger 2004; Drummond et al. 2011; for math see Cavalluzzo et al. 2012; Pane, Griffin, McCaffrey, and Karam 2013; for both see Dynarski et al. 2007). Exceptions include both strong positive and strong negative effects (for positive effects see He, Lindon, and MacLeod 2008; Banerjee, Cole, Duflo, and Linden 2009; Barrow, Markman, and Rouse 2009; for negative effects see Lindon 2008; Pane, McCaffrey, Slaughter, Steele, and Ikemoto 2010).[13]

These generally null *average* test-score effects may, however, be masking important differences from classroom to classroom. For example, Barrow, Markman, and Rouse (2009) show that the test-score gains from CAI are larger for students who should benefit most from an individualized pace of instruction: students in large classes, students far behind their peers academically, and students with poor school attendance rates. I focus in this paper on differences between teachers in how CAI affects their productivity; a question as yet

[13] Results cited in this paragraph track outcomes for just one school year: the teacher’s first year using the software. Outcomes in the second year are occasionally measured, but just as mixed (Campuzano et al. 2009, Pane et al. 2013).

unaddressed in the literature on computers in schools or teacher productivity generally.

In this study I use data from four of the experiments cited above: Dynarski et al. (2007), Barrow, Markman, and Rouse (2009), Drummond et al. (2011), and Cavalluzzo et al. (2012). All but two of the 18 products tested had no effect (or at least null results) on average test scores.[14] None of the four original analyses examined how CAI affects teacher productivity.

2. Setting, data, and experimental designs

Data for this study were collected in four field-experiments conducted during the past decade. In each experiment, teachers randomly assigned to the treatment condition received computer-aided instruction (CAI) software to begin using in their classrooms. As described earlier, in nearly all cases, the treatment had no detectable effect on average student test scores (Appendix Table A1). Table 1 summarizes the key details of each experiment: randomization design, products tested, grade-levels and subjects, and key data collected.

Collectively the experiments tested 18 different CAI software programs in reading classes in grades 1, 4, and 6; and mathematics in grade 6, pre-algebra (typically grade 8), and algebra (typically grade 9). The combined analysis sample includes more than 650 teachers and 17,000 students in over 200 schools and 80 districts from all regions of United States.[15] By design, participating schools generally had low levels of student achievement at baseline and served mostly students in poverty. Table 2 reports statistics for available student and teacher

[14] Appendix Table A1 reports mean test-score effects both from the original study reports (Column 1) and from my own re-analysis (Columns 2 and 4). The two exceptions to null effects are: Barrow, Markman, and Rouse (2009) who find a positive effect for ICL of 0.17 student standard deviations; I find the same result. Cavalluzzo et al. (2012) report a non-significant but negative effect of 0.15 student standard deviations; I find essentially the same negative point estimate, but estimate it with sufficient precision to find it statistically significant.

[15] Three of four experiments were funded by the Institute for Education Sciences, U.S. Department of Education. IES requires that all references to sample sizes be rounded to the nearest 10.

characteristics. Many schools were in large urban districts, but suburban and rural schools also participated. All schools and districts volunteered to participate.

Data from classroom observations shows strong take-up of the treatment, at least on the extensive margin: students were observed using CAI in 79 percent of math teachers' classes and 96 percent of reading teachers' classes (see Table 5 Row 1). In control classrooms the rates were 15 and 17 percent respectively. Throughout the paper I limit discussion to intent to treat effects in the interest of space. Two-thirds of treatment classes used CAI on computers in their own classroom, and one-third in shared computer labs (data from EET study only). The experiments lasted for one school year, thus, all outcomes were measured during teachers' first year using the new software.

2.A Data

Students in all experiments were tested both pre- and post-experiment using standardized achievement tests. Starting at scale score units, I standardize (mean 0, standard deviation 1) all test scores within cells defined by grade, subject, and test form, using control means and standard deviations. Test publisher and form varied by grade, subject, and experiment (see Table 1); but all tests were "low stakes" in the sense that students' scores were not used for formal accountability systems like No Child Left Behind or teacher evaluation.[16] Each experiment also collected some, but differing, demographic characteristics of students and teachers.

Three of the four experiments conducted classroom observations to measure how teachers divided class time among different tasks and activities.[17] Using these data, I measure the proportion of class time spent in three categories: whole class instruction or lectures, small group activities, and individual student

---

[16] In all but one case, the NROC experiment, the tests were administered only for purposes of the experiment.

[17] Appendix C describes the differences in data collection, and my decisions in combining data.

work. The observation data also record whether CAI software—either study-provided products or other products—was used during the class.

One experiment, the "Evaluation of Educational Technologies" (EET), also conducted extensive interviews with teachers twice during the study school year. Most notably, in the spring interviewers asked teachers to estimate how many hours, in or out of school, they spent in a typical week on various work-related tasks: teaching, preparing lessons, grading, and administering tests. I use teachers' responses to examine labor-leisure decisions. For treatment teachers, the EET interviews also include several questions about CAI use specifically: time spent learning the software, adjusting lesson plans, and setting up the systems; frequency of technical problems; use of software reports provided by the software; and others.

2.B Experimental designs

All four studies divided teachers between treatment and control conditions by random assignment, but with somewhat different designs. The "Evaluation of Educational Technologies" (EET) study, which included 15 different CAI products, and the evaluation of *Thinking Reader* (TR) both randomly assigned teachers within schools. In the EET study, all treatment teachers in the same school and grade-level were given the same CAI software product to use. The evaluation of *National Repository of Online Courses Algebra I* (NROC) randomly assigned schools within three strata defined by when the school was recruited to participate. The evaluation of *I CAN Learn* (ICL) randomly assigned classes within strata defined by class period (i.e., when during the daily schedule the class met). About one-half of teachers in the ICL experiment taught both a treatment and control class.

To assess whether the random assignment procedures were successful, I compare the average pre-treatment characteristics of treatment and control samples in Table 2. The samples are relatively well balanced, though observable characteristics differ from experiment to experiment. Both treatment teachers and students appear more likely to be male, but I cannot reject a test of the joint null of all mean differences equal to zero. Additional details on random assignment procedures and additional tests are provided in the original study reports.[18]

My measurement of teacher productivity requires student observations with both pre- and post-experiment test scores. Thus, even if samples were balanced at baseline, treatment-induced differences in attrition over the school year could bias my estimates. Since, as I describe shortly, teacher productivity is measured with student test score growth, attrition correlated with baseline test scores is of particular concern. As shown in Table 3, there is little evidence of differential attrition patterns in math classes.[19] Treatment did not affect average student (top panel) and teacher (bottom panel) attrition rates, nor did treatment change the relationships between baseline test scores and the likelihood of attrition. In reading classes, however, treatment appears to have reduced attrition overall, but increased the likelihood that a teacher would attrit if assigned a more heterogeneously skilled class. As shown in the appendix, these reading attrition differences are largely limited to the TR experiment. Notably, though, attrition rates for teachers were very low in both subjects—less than two percent of all teachers attrited.

3. Effects of CAI on the variance of teacher productivity

My first empirical objective is to estimate the causal effect of treatment—providing new CAI technology to classroom teachers—on the *variance* of teacher productivity. Throughout the paper I focus on one aspect of productivity: a

---

[18] Results of an additional test are also consistent with random assignment: I apply the methods described in Section 3.A but replace outcome test score with baseline test score. *p*-values for this test range between 0.53-0.60.

[19] Within panels, each column in Table 3 reports coefficients from a linear probability model with "attrited" as the outcome. All models include fixed effects for randomization blocks.

teacher's contribution to student academic achievement as measured by test score growth. A large literature documents substantial variability in this measure of productivity (Jackson, Rockoff, and Staiger 2014), and recent evidence suggests that variability is predictive of teacher productivity differences measured with students' long-run economic and social outcomes (Chetty, Friedman, and Rockoff 2014b).

3.A Methods

A teacher's contribution to her students' test scores is not directly observable. To isolate the teacher's contribution, I assume a statistical model of student test scores where a test score, $A_{i,t}$, for student i at the end of school year t can be written

$$A_{i,t} = f_{e(i)}\left(A_{i,t-1}\right) + \psi_{s(i,t)} + \mu_{j(i,t)} + \varepsilon_{i,t}. \tag{1}$$

The $\mu_{j(i,t)}$ term represents the effect of teacher j on student i's test score; net of prior achievement, $f_{e(i)}\left(A_{i,t-1}\right)$, and school effects, $\psi_{s(i,t)}$. The specification in 1, now commonplace in the literature on teachers, is motivated by a dynamic model of education production, suggested by Todd and Wolpin (2003), in which prior test score, $A_{i,t-1}$, is a sufficient statistic for differences in prior inputs.

With the model in 1 as a key building block, I take two separate approaches to estimating the effect of treatment on the variance of teacher productivity

$$\delta \equiv \ var(\mu|\ T = 1) - var(\mu|\ T = 0).$$

The first approach is a least-squares estimate of the conditional variance function. Specifically, I estimate the treatment effect on the variance $\hat{\delta}^{LS}$ by fitting

$$\left(\mu_j - \mathbb{E}\left[\mu_j|\ T_j, \pi_{b(j)}\right]\right)^2 \ = \delta^{LS} T_j + \pi_{b(j)} + \nu_j, \tag{2}$$

where T_j is an indicator $= 1$ if the teacher was assigned to the CAI treatment and zero otherwise, and $\pi_{b(j)}$ represent fixed effects for each randomization block group, b. The latter are included to account for the differing probabilities of selection into treatment; probabilities dictated by each experiment's design (i.e., random assignment within schools, recruitment strata, or class period).

My approach to estimating Specification 2 has three steps. Step one, estimate μ, as described in the next paragraph. Then follow the common, feasible approach to fitting conditional-variance specifications like 2: Step two, estimate $\mathbb{E}[\hat{\mu}_j | T_j, \pi_{b(j)}]$ by ordinary least-squares, i.e., fit $\hat{\mu}_j = \tilde{\delta} T_j + \tilde{\pi}_{b(j)} + u_j$.[20] Step three, estimate Specification 2 using the squared residual from step two, $\hat{u}_j^2$, as the dependent variable. I calculate standard errors for $\hat{\delta}^{LS}$ that allow for clustering within schools.

In step one I estimate the test-score productivity of each teacher, $\hat{\mu}_j$, by fitting Equation 1 treating the $\mu_{j(i,t)}$ as teacher fixed effects.[21,22] The $\psi_{s(i,t)}$ terms are school fixed effects, and $f_{e(i)}$ is a quadratic in pre-experiment test score. The parameters of $f_{e(i)}$ are allowed to be different for each of the various tests, e, used to measure $A_{i,t}$ and $A_{i,t-1}$; each e is defined by the intersection of grade-level, subject, and experiment. Note that this teacher-fixed-effects approach does not require a distributional assumption about $\mu_{j(i,t)}$, and identifies other model parameters using only within-teacher variation. Finally, the estimated teacher fixed effects, $\hat{\mu}_j$, include estimation error. I "shrink" the $\hat{\mu}_j$, multiplying each

[20] It may seem intuitive to interpret the estimate of $\tilde{\delta}$ from step two as the treatment effect on the mean of teacher productivity. However, as I discuss further in Section 4, the mean productivity effect cannot be separately identified.

[21] The teacher fixed effects are parameterized to be deviations from the school average, rather than deviations from an arbitrary hold out teacher, using the approach suggested by Mihaly, McCaffrey, Lockwood, and Sass (2010).

[22] Kane and Staiger (2008) and Chetty, Friedman, and Rockoff (2014a) use an alternative approach to estimating $\hat{\mu}_j$ which, in short, uses average test-score residuals. My estimates of $\hat{\delta}^{LS}$ are robust to taking this alternative approach.

estimate by its estimated signal-to-total variance ratio.[23,24]

The second approach to estimating δ is a maximum likelihood estimate, $\hat{\delta}^{ML}$, obtained by treating $\mu_{j(i,t)}$ as teacher random effects. I fit a slightly re-parameterized version of Equation 1,

$$A_{i,t} = f_{e(i)}\left(A_{i,t-1}\right) + \psi_{s(i,t)} + \mu^{T}_{j(i,t)}T_{j(i,t)} + \mu^{C}_{j(i,t)}\left(1 - T_{j(i,t)}\right) + \varepsilon_{i,t}, \tag{3}$$

where $\mu^{T}_{j(i.t)}$ and $\mu^{C}_{j(i.t)}$ are random effects with the assumed distribution

$$\begin{bmatrix}\mu^{T}_{j(i,t)}\\ \mu^{C}_{j(i,t)}\end{bmatrix} \sim N(\begin{bmatrix}\mu^{T}\\ \mu^{C}\end{bmatrix}, \begin{bmatrix}\sigma^{2}_{\mu^{T}} & 0\\ 0 & \sigma^{2}_{\mu^{C}}\end{bmatrix}).$$

That is, the model allows the estimated variance of the teacher-specific random intercepts to differ between treatment and control. I also allow the variance $\varepsilon_{i,t}$ to be difference for treatment and control groups. As in the least-squares approach, $\psi_{s(i,t)}$ are school fixed effects and $f_{e(i)}$ is a quadratic function specific to each test. Maximum likelihood estimation of this linear mixed model provides $\hat{\sigma}^{2}_{\mu^{T}}$ and $\hat{\sigma}^{2}_{\mu^{C}}$, and thus $\hat{\delta}^{ML} = \left(\hat{\sigma}^{2}_{\mu^{T}} - \hat{\sigma}^{2}_{\mu^{C}}\right)$.

To interpret either of the two estimates, $\hat{\delta}^{ML}$ or $\hat{\delta}^{LS}$, as the causal effect of new CAI software on the variance of teacher productivity requires two assumptions. *Assumption 1:* At the start of the experiment, there was no difference between the treatment teachers and control teachers in teachers' potential for making productivity gains (losses) during the experiment school year. This assumption should be satisfied by the random assignment study designs.

[23] I estimate signal variance with the total variance of $\hat{\mu}_j$ minus the mean squared standard error of the $\hat{\mu}_j$. Signal variance is estimated separately for treatment and control samples. The total variance for estimate *j* is signal variance plus the squared standard error of $\hat{\mu}_j$.

[24] Estimates of the treatment effect on the variance are still statistically significant, and not substantially different if I do not shrink the teacher fixed effect estimates.

Assumption 2: Students were not assigned to teachers based on unobserved (i.e., omitted from Equation 1 or 3) determinants of potential for test score growth: $\mathbb{E}\left[\varepsilon_{i,t} \mid j\right] = \mathbb{E}\left[\varepsilon_{i,t}\right]$. This assumption is necessary for obtaining consistent estimates of $\hat{\mu}_j$, and parameters like it throughout the teacher effects literature. Empirical tests of this assumption by Chetty, Friedman, and Rockoff (2014a) and Kane and Staiger (2008) find little residual bias in $\hat{\mu}_j$ if the estimating equation includes, as I do, flexible controls for students' prior achievement, and controls for teacher and student sorting between schools.[25,26]

Assumption 2 is, strictly speaking, only needed to identify the *levels* of variance. A weaker alternative is sufficient for causal estimates of the *relative difference* in variance, and thus the sign of $\hat{\delta}^{ML}$ or $\hat{\delta}^{LS}$. *Assumption 2 Alternative:* Any source of (residual) bias in estimating $\hat{\mu}_j$ is independent of the condition, treatment or control, to which a teacher was assigned. Like Assumption 1, this alternative assumption should be satisfied by the random assignment of teachers.

One final note about methods, the experiment for *I CAN Learn* randomly assigned classes, not teachers, to treatment and control conditions. Half of teachers in that experiment taught both a treatment and control class. Except where explicitly noted in one analysis, I treat each ICL class as a separate observation j and estimate a separate μ_j. I show the results are robust to excluding ICL entirely from the estimation sample. Moreover, the inclusion of ICL appears to attenuate the estimated effect of treatment on the variance of productivity (see Appendix Table A2). The smaller effects in the ICL sample may be the result of

---

[25] For detailed discussions of the theoretical and econometric issues in isolating teacher contributions to student test score growth see Todd and Wolpin (2003), Kane and Staiger (2008), Rothstein (2010), and Chetty, Friedman, and Rockoff (2014a).

[26] Students were not randomly assigned to classes or teachers in any of the four experiments. Schools in the ICL experiment claimed the class assignment process, carried out by software, close to random; and tests of the data are consistent with that claim (Barrow, Markman, and Rouse 2009, footnote 9).

teachers with both types of class re-allocating saved effort from their treatment class to their control class.

3.B Estimates

At least in (middle- and high-school) math classes, providing teachers with computer-aided instruction software for use in their classrooms substantially reduces the variability of teacher productivity, as measured by student test score growth. Columns 1 and 4 of Table 4 report the estimated standard deviation of teacher productivity in the control group, measured in student standard deviation units. Columns 2 and 5 report estimates of $\hat{\delta}^{LS}$ and $\hat{\delta}^{ML}$, respectively, using the pooled sample of all experiments. In treatment mathematics classes, the standard deviation of teacher productivity fell by between one-quarter and one-half. This change is consistent with the prediction that labor-replacing technology should reduce the variation in teacher productivity. In (elementary and middle-school) reading classes, by contrast, there was no statistically significant or practical difference.

The treatment effects in math are educationally substantial. In control classrooms, students assigned to a teacher at the 75th per centile of the job performance distribution will score approximately 0.20 standard deviations higher on achievement tests than their peers assigned to the median teacher. (The estimated control standard deviation is on the high end of existing estimates, see Jackson, Rockoff, and Staiger 2014.) By contrast, in treatment classrooms a student's teacher assignment has become much less consequential. The median to 75th percentile difference is just 0.12 to 0.15 standard deviations.

This reduction in variance is partly due to the standardization that intuitively occurs when using a computer to carry out some task(s). But the magnitude of the reduction is also partly due to changes in how teachers' choose to carry out their work day to day—changes induced by the option of using CAI. I discuss those changes in Section 5.

As shown in Appendix Table A2, the main pattern of results in Table 4 is not driven by a particular CAI product, nor the data from a particular experiment. In Appendix Table A2 I repeat the entire estimation process on subsamples which iteratively exclude one experiment at a time. The robustness of the estimates across samples is strong evidence that the treatment effects are a general characteristic of computer-aided instruction rather than the idiosyncratic characteristic of one particular experiment or software program. Additionally, the robustness across experimental designs suggests spillover effects were limited—across treatment and control teachers in the same school, or across treatment and control classes taught by the same teacher—since the experiments' different designs each permitted different opportunities for spillovers.

The remaining two rows in Appendix Table A2 test sensitivity to the omission of school fixed effects (in Equations 1 and 3). For some purposes estimates of $\hat{\mu}_j$ and $\hat{\sigma}_{\mu}^2$ without school fixed effects are preferred, and, given the alternative assumption 2, the inclusion or exclusion should not dramatically affect the inferences of interest. In place of school fixed effects I include fixed effects for the randomization block groups, b. The results are very robust to this change.

Finally, the absence of effects in reading classes is striking next to the large effects in math classes. As I show later, reading teachers were equally likely to use the software, and made similarly large changes in their use of classroom time. I raise two possible explanations for the lack of effects in reading. First, CAI may replace teachers' labor in aspects of reading instruction where teachers' contributions are already (relatively) homogeneous. Researchers have long noted that teachers' estimated effects on reading test scores vary less than their estimated effects on math scores (Jackson, Rockoff, and Staiger 2014). Second, alternatively, CAI may replace teachers' labor in aspects of reading instruction that the typical standardized reading test does not measure. Kane and Staiger (2012) show that differences in teachers' observed instructional skills do predict differences in student test scores on an atypical reading test which measures a broad range of reading and writing skills, but those same observed skills do not predict differences on a typical narrowly-focused standardized reading test.

4. Heterogeneity of effects on teacher productivity

If CAI reduces variation in teacher productivity, a critical follow-up question is whether the reduction is the result of productivity improvements among otherwise low-performing teachers, or productivity losses among otherwise high-performing, or both. The variance could also shrink if the productivity of all teachers improved (declined), but the relatively low-performing teachers improved more (declined less). In this section I test whether the treatment effects on teachers are heterogeneous, in particular whether the CAI-induced change in a teacher's productivity is related to her counterfactual productivity level.

4.A Methods

To test for treatment effect heterogeneity I examine the quantiles of $\hat{\mu}_j$—the teacher productivity estimates described in Section 3—comparing quantiles of the treatment teacher distribution to quantiles of the control distribution. Recall that the τth quantile of the cdf $F(y)$, denoted $q_\tau(y)$, is defined as the minimum value of y such that $F(y) = \tau$.

I begin by simply plotting the quantiles of $\hat{\mu}_j$ separately for treatment (solid line) and control teachers (dotted line) in Figure 2. These plots are traditional cdf plots with the axes reversed. Each line traces out a series of quantiles calculated at increments of $\tau = 0.01$, for example, $q_{\tau=0.01}(\hat{\mu}_j | T_j = 1) \ldots q_{\tau=0.99}(\hat{\mu}_j | T_j = 1)$ for the solid line. In Figure 2, and throughout this section, I show results using teacher fixed effects estimates of $\hat{\mu}_j$ obtained as described in Section 3. Results using, instead, best linear unbiased predictions

(BLUPs) of teacher random effects show similar patterns.

Our interest is in the vertical distance between the solid and dotted lines: the difference between the productivity level of a τth percentile treatment teacher and a τth percentile control teacher. To obtain point estimates for these vertical differences, $\hat{\gamma}_\tau$, I use the unconditional quantile regression method proposed by Firpo, Fortin, and Lemieux (2009). The regression specification is

$$[\tau - \mathbf{1}\{\mu_j \leq q_\tau\}]\left[f_{\mu_j}(q_\tau)\right]^{-1} = \gamma_\tau T_j + \pi_{\tau,b(j)} + \epsilon_{\tau,j}, \tag{4}$$

where the dependent variable on the left is the influence function for the τth quantile, $\text{IF}\left(\mu_j; q_\tau, F_{\mu_j}\right)$, T_j is the treatment indicator, and $\pi_{\tau,b(j)}$ the randomization block fixed effects. Firpo and colleagues detail the properties of this IF-based estimator, which are straightforward in this randomly-assigned binary treatment case.[27] For inference I use cluster-bootstrap standard errors (500 replications) which allow for dependence within schools. The interpretation of the quantile treatment effects $\hat{\gamma}_\tau$, and their relevance to effect heterogeneity, involves some subtleties which I take up below, but the basic causal warrant rests on Assumptions 1 and 2 described in Section 3.

Before creating Figure 2 or estimating γ_τ, I make one modification to the teacher productivity estimates: I set the mean of $\hat{\mu}_j$ to zero within each CAI-product-by-treatment-condition cell. The motivation is that *mean* teacher productivity effects are not identified separately from the total mean effects of treatment. In practical terms, a treatment indicator would be collinear with teacher fixed effects when estimating Equation 1. If there are large positive (negative)

[27] Firpo (2007) develops an alternative approach to estimating unconditional quantile treatment effects using propensity score weighting. Perhaps not surprisingly, the results presented in Figure 3 are robust to taking this alternative approach. In this setting, the Firpo (2007) approach simplifies to calculating $\hat{\gamma}_\tau = q_\tau(\mu_j|T_j = 1) - q_\tau(\mu_j|T_j = 0)$ where each observation j is weighted by the inverse probability of treatment (IPTW).

average effects on teacher productivity, then de-meaning $\hat{\mu}_j$ would induce negative (positive) bias in estimates of γ_τ. However, the total mean treatment effect estimates in Appendix Table A1 are almost all null suggesting such bias is not a first-order concern.

4.B Estimates

The treatment-control differences in the top-panel of Figure 2 suggest computer-aided instruction software can affect the productivity of different math teachers in quite different ways. The use of CAI appears to improve the productivity of otherwise low-performing math teachers, yet simultaneously lower the productivity of otherwise high-performing teachers. In contrast to math, but consistent with the results in Table 4, there is little if any difference for reading teachers. In Figure 3, focusing on math teachers, I plot the estimated unconditional quantile treatment effects $\hat{\gamma}_\tau$ and 95 percent confidence intervals.

The interpretation of these treatment-control differences requires some subtlety. There are two interpretations. First, without any further assumptions, we can use Figure 2 and the estimates $\hat{\gamma}_\tau$ to describe changes in the *distribution* of teacher productivity brought on by the introduction of CAI software. Imagine two schools, identical except that school A uses CAI and school B does not. The estimates in Figure 3 suggest that in school A the impact of being assigned to a bottom-quartile teacher instead of a top-quartile teacher will be much less consequential than in school B. But this reduction in the consequences of teacher assignment comes partly because students in the classrooms of school A's top teachers are not learning as much as students in the classrooms of school B's top teachers.

More generally, for each quantile τ, $\hat{\gamma}_\tau$ measures the difference in the two productivity distributions at the τth percentile, for example, the "difference in median productivity" when $\tau = 0.5$. Thus the series of $\hat{\gamma}_\tau$ in Figure 3 provide an alternative description of how treatment affects the variability in teacher productivity—less-parametric than the estimates in Table 4 but at the cost of less precision.

As the language of the school A versus school B example indicates, this first interpretation has clear relevance to management and policy decisions. In particular, this first interpretation is relevant when considering CAI as an intervention alongside other interventions, like more-selective hiring and firing or on-the-job training, aimed at improving the stock of teaching quality generally.

A second, though not mutually exclusive, interpretation is that $\hat{\gamma}_\tau$ measures the causal treatment effect of CAI on teachers' at the τth percentile of the teacher productivity distribution. Under this interpretation, for example, the estimates in Figure 3 suggest CAI cuts productivity by 0.07 student standard deviations among 75th percentile teachers, but raises productivity by 0.09 among 25th percentile teachers. This second interpretation also has value for management of teachers, particularly the supervision of *individuals*. Heterogeneous effects may prompt school principals to encourage (permit) CAI use by some teachers but not others.

This second interpretation requires a third assumption of rank-invariance. *Assumption 3*: While treatment may have changed productivity *levels*, it did not change the *rank ordering* of teachers in terms of estimated productivity. This third assumption is unlikely to hold perfectly. However, even if this assumption is violated, we can still make some causal conclusions about treatment effect heterogeneity from Figure 3. Specifically, if the estimated treatment effect at one point in the distribution is positive (negative), then treatment improved (lowered) productivity for at least some teachers (Bitler, Gelbach, and Hoynes 2003).[28]

[28] A different question of heterogeneity is whether the *mean* effect on test scores for a given CAI software covaries with that software's effect on the *variance* of teacher performance. To test this question I estimated (i) total mean effect and (ii) effect on the variance of teacher effects by each

5. Effects of CAI on teachers' instructional choices and effort

In this final section I estimate the effects of treatment—computer-aided instruction software—on teachers' decisions about how to allocate classroom time across different activities, and on teachers' level of work effort. Changes in teachers' decisions and effort are potential mechanisms behind the estimated changes in teacher productivity, especially mechanisms relevant to magnitude and heterogeneity of treatment effects.

5.A Effects on teachers' allocation of class time

I first examine whether computer-aided instruction software changes how teachers divide class time among different tasks or activities. In three of four experiments researchers observed teachers and students during class time, and at regular intervals recorded what instructional activities were taking place: (i) lecturing or whole-class activities, (ii) students working individually, (iii) students working in pairs or small groups.[29] In Table 5, I report the proportion of class time control teachers allocated to each of these three tasks, on average, and the treatment-control differences in time allocation. Each reported treatment effect on the proportion of class time, $\hat{\beta}$, is estimated in a simple least-squares regression

$$y_j = \beta T_j + \pi_{b(j)} + \eta_j, \tag{5}$$

---

of the 18 CAI programs, and then correlated (i) and (ii). The results should be interpreted with caution given the small sample for any one program, and thus large standard errors on (i) and (ii). In math the larger is the mean effect the smaller is the reduction in between-teacher variance. There is apparently no relationship in reading.

[29] The observation protocols and data collection instruments differed somewhat across studies. The primary differences were in the level of detail collected; for example, recording "small groups" and "pairs" as two separate activities versus one activity, or recording data at 7 minute intervals versus 10 minute intervals. Appendix C describes the differences in protocols and instruments, and my decisions in combining the data. The pattern of results in Table 5 holds for each of the three studies when analyzed individually (Appendix Table C2).

which includes the same treatment indicator, T_j, and randomization block fixed effects, $\pi_{b(j)}$, as used in earlier sections of the paper. Standard errors allow for clustering within schools. $\hat{\beta}$ is identified by the random assignment designs.

Data from direct observations in classrooms show notable changes in teachers' instructional choices and practices. Treatment teachers doubled the share of class time devoted to students working individually, on average; the added individual time would have otherwise been devoted to lectures or other whole-class activities. As reported in Table 5, this pattern is true of both math and reading classrooms. In math classrooms the share of class time allocated to individual student work increased from 38 percent of class time to 73 percent. Simultaneously, the share of time in whole-class activities fell by half, from 61 percent to 30 percent.[30] The magnitudes are similar in the reading classes.[31]

This reallocation of class time, from lectures to individual work, is consistent with teachers who are making rational production decisions—responding to changes in the marginal productivity and marginal costs of individual student time. This interpretation assumes that the changes in productivity and costs do, on average, favor increasing the use of individual student work. That would be true in the plausible case, described in Appendix B, where using CAI increases the marginal productivity of time allocated to individual student work and simultaneously decreases the marginal costs. Theoretical reasons to expect these two conditions with CAI are discussed in

---

[30] Teachers could allocate class time to multiple tasks simultaneously. As shown in Table 5, researchers observed multiple activities at once about 20 percent of the time in math classes, and about 25 percent of the time in reading classes; accordingly the average allocations do not sum to one. However, treatment did not affect the frequency of multi-tasking.

[31] Additionally, reading teachers (in self-contained 1st and 4th grade classes) also increased the total amount of time spent on reading instruction, presumably at the expense of other subjects like math or art. The treatment effect estimate is shown in Appendix Table A3 Row 2; the data and estimation are described in Section 5.B. Indeed, total time spent on reading in a typical week roughly doubled. The increase in reading instruction may help explain why there was little effect on the variance of reading teacher productivity.

Appendix B. Empirical tests of these conditions are limited by scarce data, but the results are consistent.

First, to test for a change in the productivity of individual student time, I regress estimated teacher fixed effects, $\hat{\mu}_j$, on the treatment indicator, T_j; the three class time measures, the y_js; and the interaction of T_j and each y_j.[32] The coefficient on "individual student work" time is, as predicted, somewhat larger for math treatment teachers than control teachers, though the difference is not statistically significant. Full results are available in the appendix.

Second, to test for changes in costs, I examine measures of teacher effort. Two indirect measures of effort during class time are shown in the bottom panel of Table 5, and the estimation methods match the rest of Table 5. The results are consistent with a reduction in teacher effort costs. Treatment teachers' most common role in class activities was "facilitating", while control teachers were most often "leading" the class activity. Managing student behavior was also apparently less of a challenge in treatment math classes. Additionally, as described in the next section and shown in Table 6, treatment math teachers also spent fewer hours outside of class time doing preparatory work like grading and planning. Finally, in data gathered only for treatment classrooms, observers reported that nearly half of teachers took no active role when students were using CAI, and that technical difficulties with the software were relatively infrequent (occurring in just 27 percent of observations).

A second interpretation of the time allocation changes is that teachers increased individual student time to comply with expectations of or recommendations by their manager, the software publishers, or the researchers, without regard to how it might affect productivity. I do not have data to test this

[32] I also include randomization block fixed effects, and report clustered (school) standard errors.

second interpretation. However, the rational-decision interpretation and compliance interpretation are not mutually exclusive.

5.B Effects on teachers' total work hours

Finally, I measure how teachers' total work hours change after they are given CAI software to use in their classrooms. In structured interviews, teachers reported how many hours, in or out of school, they spent during a typical week on various work-related tasks for one typical class: teaching students, preparing lessons, grading, and administering tests. (These data are only available for the EET study.) Table 6 shows control teachers' reported hours in each task, on average, and the percentage difference between treatment and control hours. The estimated treatment effects come from fitting Specification 5 with log hours as the dependent variable.

Among math teachers, the software reduces work effort on the extensive margin. As reported in Table 6, treatment math teachers worked 23.4 percent fewer hours than their control colleagues. Time in the classroom did not change; both treatment and control teachers report about three hours per week teaching students, or about 35 minutes per day for the typical math class.[33] But treatment math teachers spent less time planning lessons and grading, about one-third fewer hours in a typical week.

This reduction in total work effort may help rationalize the behavior of math teachers who choose to use CAI in their classrooms despite the reduction in their productivity. (Of course, the reduction in total effort only reinforces the adoption of CAI for teachers whose productivity improves.) In short, a teacher may rationally trade smaller student achievement gains for reduced work effort.

[33] These self-reported data on work hours are vulnerable to important sources of measurement error, even non-classical error. Accordingly, readers should be cautious about interpreting the levels, e.g., the control means. Both treatment and control teachers may have under or over reported the quantity of hours. However, the estimated treatment-control differences, $\hat{\beta}$, in Table 6 are nevertheless interpretable as causal effects as long as any source of reporting bias or error is independent of treatment assignment.

Unfortunately, the data in this study do not permit a thorough analysis of the relationship between changes in productivity and changes in effort. [34]

The student achievement losses among otherwise high-performing teachers, seen in Figure 2, may have an explanation outside the simple model. Perhaps high-performers were maximizing inter-temporally, and viewed the first year as a training investment; or high-performers felt an obligation to their managers or the research project even if their students were made worse off. Both examples suggest high-performing treatment teachers would be working harder during the experiment year in order to ameliorate the student achievement losses that would come with using CAI. In particular, a teacher should inter-temporally smooth utility, to some extent, by increasing effort in the first year using CAI. The estimates showing reduced hours in Table 6 Column 2 are average effects pooling all teachers; the averages may be driven by reductions among the low-performers masking hours increases for high-performers.

However, if high-performing teachers were making an achievement-effort trade off, we would instead expect to see reduced hours among (at least some) high-performers. To test this prediction I, first, divide teachers into terciles of $\hat{\mu}_j$, and then, second, estimate changes in work hours separately for each tercile.[35] The results are reported in Table 6 Columns 3-5. Consistent with the prediction, relatively high-performing teachers (top tercile) did reduce their total work hours, and indeed that reduction was larger than that of their relatively low-performing colleagues though the difference is not statistically significant.

[34] Teachers did report various fixed costs of using CAI. Treatment math teachers in the EET study reported spending, on average, 3.1 hours (s.d. 3.7) learning the software, 2.1 hours (3.4) setting up and configuring the software, and 4.1 hours (8.9) updating lesson plans. Fifty-nine percent said that they were given sufficient paid time to accomplish these tasks. These fixed costs, while positive, are small compared to the estimated recurring effort savings.

[35] The results in Table 6 Columns 3-5 come from a single regression, identical to Specification 5 except that I interact T_j with indicators for terciles of teacher productivity. The terciles of $\hat{\mu}_j$ are determined separately for treatment and control distributions.

This approach is non-standard, notably because the productivity terciles are based on post-treatment outcomes. Identification requires an assumption akin to rank invariance, but somewhat weaker: while treatment may have changed productivity, it did not change the productivity tercile to which a teacher belongs.

6. Conclusion

Differences in teachers' access to and use of technology—like computer-aided instruction (CAI) software—contribute to differences in teachers' productivity and professional practices. Providing math teachers CAI for use in their classrooms substantially shrinks the variance of teacher productivity, as measured by teacher contributions to student test score growth. The smaller variance comes from both productivity improvements among otherwise low-performing teachers, but also productivity losses among some high-performing teachers. These changes in productivity partly reflect technology-induced changes in how teachers choose to accomplish their work: technology affects both teachers' work effort and their decisions about how to allocate class time. These results are some of the first empirical evidence on how new technology affects teacher productivity.

The analysis in this paper highlights, more broadly, how both teachers' skills and teachers' decisions contribute to their observed productivity. Replacing teacher labor with machines, like the computer-aided instruction example I examine, can greatly benefit students in some classrooms, especially the classrooms of low performing teachers, while simultaneously making students in other classrooms worse off. This difference in outcomes arises partly because, given the option, some teachers may choose to use a new technology, even if it reduces their students' achievement, because it also substantially reduces their workload. Understanding teachers' work decisions is critical to better research, and to better management and policy decisions.

References

Acemoglu, D. & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics, 4*, 1043-1171.

Acemoglu, D., Laibson, D., & List, J. A. (2014). Equalizing superstars: The internet and the democratization of education. NBER 19851.

Angrist, J. & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *The Economic Journal, 112*(482), 735-765.

Atkin, D., Chaudhry, A., Chaudry, S., Khandelwal, A. K., & Verhoogen, E. (2017). Organizational barriers to technology adoption: Evidence from soccer-ball produces in Pakistan. *Quarterly Journal of Economics, 132*(3), 1101-1164.

Autor, D., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4), 1279-1333.

Autor, D., Katz, L. F., & Krueger, A. B. (1998). Computing inequality: Have computers changed the labor market? *Quarterly Journal of Economics*, *113*(4), 1169-1213.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, *122*(3), 1235-1264.

Barrera-Osorio, F. & Linden, L. L. (2009). The use and misuse of computers in education: Evidence from a randomized controlled trial of a language arts program. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (J-PAL).

Barrow, L., Markman, L., & Rouse, C. E. (2008). Technology's edge: The educational benefits of computer-aided instruction. NBER 14240.

Barrow, L., Markman, L., & Rouse, C. E. (2009). Technology's edge: The educational benefits of computer-aided instruction. *American Economic Journal: Economic Policy*, *1*(1), 52-74.

Barlevy, G. & Neal, D. (2011). Pay for percentile. *American Economic Review*, *102*(5), 1805-1831.

Betts, J. R. & Shkolnik, J. L. (1999). The behavioral effects of variations in class size: The case of math teachers. *Educational Evaluation and Policy Analysis*, *21*(2), 193-213.

Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *The American Economic Review*, *96*(4), 988-1012.

Brown, B. W. & Saks, D. H. (1987). The microeconomics of the allocation of teachers' time and student learning. *Economics of Education Review*, *6*(4), 319-332.

Campuzano, L., Dynarski, M., Agodini, R., and Rall, K. (2009). Effectiveness of reading and mathematics software products: Findings from two student cohorts. Washington, DC: U.S. Department of Education.

Cavalluzzo, L., Lowther, D., Mokher, C., and Fan, X. (2012). Effects of the Kentucky Virtual Schools' hybrid program for algebra I on grade 9 student math achievement. Washington, DC: U.S. Department of Education.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593-2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and outcomes in adulthood. *American Economic Review, 104*(9), 2633-2679.

Cuban, L. (2001). *Oversold and underused: Computers in schools 1980-2000.* Cambridge, MA: Harvard University Press.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, *37*(4), 696-727.

Drummond, K., Chinen, M., Duncan, T.G., Miller, H.R., Fryer, L., Zmach, C., & Culp, K. (2011). Impact of the Thinking Reader software program on grade 6 reading vocabulary, comprehension, strategies, and motivation. Washington, DC: U.S. Department of Education.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H., Emery, D., & Sussex, W. (2007). Effectiveness of reading and mathematics software products: Findings from the first student cohort. Washington, DC: U.S. Department of Education.

Education Week. (2013, January 17). Market for education software, digital products has grown, analysis shows. *Education Week.* Retrieved from http://blogs.edweek.org/edweek/marketplacek12/2013/01/

Fairlie, R. W. (2005). The effects of home computers on school enrollment. *Economics of Education Review, 24*(5), 533-547.

Fairlie, R. W., Beltran, D. O., & Das, K. K. (2010). Home computers and education outcomes: Evidence from the NLSY97 and CPS. *Economic Inquiry, 48*(3), 771-792.

Fairlie, R. W. & London, R. A., (2012). The effects of home computers on education outcomes: Evidence from a field experiment with community college students. *Economic Journal 122*(561), 727-753.

Fairlie, R. W. & Robinson, J. (2013). Experimental evidence on the effects of home computers on academic achievement among schoolchildren. *American Economic Journal: Applied Economics, 5*(3), 211-240.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects, *Econometrica*, *75*(1), 259-276.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953-973.

Fuchs, T. & Woessmann, L. (2004). Computers and student learning: Bivariate and multivariate evidence on the availability and use of computers at home and at school. CESifo working paper 1321.

Gilpin, G. & Kaganovich, M. (2012). The quantity and quality of teachers: Dynamics of the trade-off. *Journal of Public Economics, 96*(3), 417-429.

Goolsbee, A. & Guryan, J. (2006). The impact of Internet subsidies in public schools. *The Review of Economics and Statistics*, *88*(2), 336-347.

He, F., Linden, L. L., & MacLeod, M. (2008). How to teach English in India: Testing the relative productivity of instruction methods within the Pratham English Language Education Program. Columbia University working paper.

Jackson, C. K. (2014). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina. NBER 18624.

Jackson, C. K., & Makarin, A. (2016). Can online off-the-shelf lessons improve student outcomes? Evidence from a field experiment. NBER 22398.

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, *6*, 801-825.

Jamison, D. T., Fletcher, J. D., Suppes, P., and Atkinson, R. C. (1976). Cost and performance of computer-assisted instruction for education of disadvantaged children. In J. T. Froomkin, D. T. Jamison, and R. Rander (Eds.), *Education as an Industry*. Cambridge, MA: National Bureau of Economic Research.

Jorgenson, D. W., Ho, M. S., & Stiroh, K. J. (2005). *Productivity, Volume 3: Information Technology and the American Growth Resurgence.* Cambridge, MA: MIT Press.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* Seattle, WA: Bill & Melinda Gates Foundation.

Kane, T. J. & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. NBER 14607.

Kane, T. J. & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation.

Kirkpatrick, H. & Cuban, L. (1998). Computers make kids smarter—right?. *Technos*, *7*(2), 26-31.

Lakdawalla, D. (2006). The economics of teacher quality. *Journal of Law and Economics*, *49*, 285-329.

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics*, *89*(4), 721-736.

Machin, S., McNally, S., & Silva, O. (2007). New technology in schools: Is there a payoff? *The Economic Journal*, *117*(522), 1145-1167.

Malamud, O. & Pop-Eleches, C. (2011). Home computer use and the development of human capital. *Quarterly Journal of Economics*, *126*(2), 987-1027.

Mihaly, K., McCaffrey, D. F., Lockwood, J. R., & Sass, T. R. (2010). Centering and reference groups for estimates of fixed effects: Modifications to felsdvreg. *Stata Journal*, *10*(1), 82-103.

Murnane, R. J. & Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research*, *10*(1), 83–100.

Murphy, R., Penuel, W., Means, B., Korbak, C., Whaley, A., & Allen, J. E. (2001). *A review of recent evidence on the effectiveness of discrete educational software.* Menlo Park, CA: SRI International.

Oliner, S., Sichel, D., & Stiroh, K. (2007). Explaining a productive decade. *Brookings Papers on Economic Activity*, 1: 81–137.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2013). Effectiveness of Cognitive Tutor Algebra I at scale. RAND working paper WR-984-DEIES.

Pane, J. F., McCaffrey, D. F., Slaughter, M. E., Steele, J. L., & Ikemoto, G. S. (2010). An experiment to evaluate the efficacy of Cognitive Tutor Geometry. *Journal of Research on Educational Effectiveness*, *3*(3), 254-281.

Pearson, P. D., Ferdig, R. E., Blomeyer Jr, R. L., & Moran, J. (2005). *The effects of technology on reading performance in the middle school grades: A meta-analysis with recommendations for policy.* Naperville, IL: Learning Point.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, *125*(1), 175-214.

Rothstein, J. (2012). Teacher quality policy when supply matters. National Bureau of Economic Research 18419.

Rouse, C. E. & Krueger, A. (2004). Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, *23*(4), 323-338.

Schmitt, J. & Wadsworth, J. (2006). Is there an impact of household computer ownership on children's education attainment in Britain? *Economics of Education Review*, *25*(6), 659-673.

Staiger, D. O. & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, *24*(3), 97-117.

Suppes, P. & Morningstar, M. (1969). Computer-assisted instruction. *Science*, *166*(903), 343-350.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature*, *49*(2), 326-365.

Taylor, E. S. & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, *102*(7), 3628-3651.

Todd, P. E. & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, *113*(485), F3-F33.

Vigdor, J. L. & Ladd, H. F. (2010). Scaling the digital divide: Home computer technology and student achievement. NBER 16078.

Table 1—Summary of experiments and CAI products tested

Grade subject	Products	Random assignment design	Student test	Teacher measures	Original report
(1)	(2)	(3)	(4)	(5)	(6)
EET—*"Evaluation of Educational Technologies"*					
1st reading	Academy of Reading Destination Reading Headsprout Plato Focus Waterford Early Reading	Teachers within schools	SAT-9	Classroom observations & teacher interviews	Dynarski et al. (2007), and Campuzano et al. (2009)
4th reading	Academy of Reading Knowledgebox Leapfrog Read 180		SAT-10		
6th math	Achieve Now iLearn Math Larson Pre-algebra		SAT-10		
Algebra	Cognitive Tutor Algebra Larson Algebra Plato Algebra		ETS EOC Algebra		
ICL—*I CAN Learn*					
Pre-algebra & Algebra	I CAN Learn	Classes within class schedule periods	NWEA		Barrow, Markman, Rouse (2009)
NROC—*National Repository of Online Courses, Algebra I*					
Algebra	National Repository of Online Courses	Schools within recruitment strata	ACT EXPLORE & ACT PLAN	Classroom observations	Cavalluzzo et al. (2012)
TR—*Thinking Reader*					
6th reading	Thinking Reader	Teachers within schools	GMRT	Classroom observations	Drummond et al. (2011)

Note: Student test abbreviations: "SAT-9" and "SAT-10" are the Stanford Achievement Test, versions 9 and 10, Pearson. "ETS EOC Algebra" is the end-of-course Algebra exam, Educational Testing Service. "NWEA" is a custom test of pre-algebra and algebra skills developed by Northwest Evaluation Association for the ICL study. "GMRT" is the Gates-MacGinitie Reading Test, Houghton Mifflin Harcourt.

Table 2—Student and teacher characteristics, and pre-treatment covariate balance

	Cont. mean (st. dev.)	Treat. mean (st. dev.)	Diff. = 0 p-value	Joint test p-value	Obs.	
	(1)	(2)	(3)	(4)	(5)	
Student characteristics				0.335		
Pre-experiment test score	0.002	-0.019	0.519		19,420	
	(0.919)	(0.925)				
Age	12.326	12.346	0.133		18,790	a
	(0.559)	(0.570)				
Female	0.501	0.489	0.021		21,810	
Minority	0.382	0.367	0.461		12,860	b
Special education program	0.101	0.117	0.154		11,540	b
Gifted and talented program	0.130	0.126	0.842		6,450	c
Free or reduced price lunch	0.730	0.773	0.199		8,310	d
Attendance rate prior year	0.825	0.829	0.703		1,060	d
	(0.164)	(0.172)				
Teacher characteristics				0.064		
Years of experience	11.134	11.412	0.630		490	e
	(6.585)	(6.445)				
Age	40.263	39.126	0.172		380	f
	(8.236)	(8.135)				
Female	0.867	0.805	0.013		420	f
Minority	0.332	0.329	0.901		380	f
Master’s degree	0.492	0.446	0.173		480	e
Regular certification	0.927	0.919	0.659		480	e
Number of students	28.121	28.338	0.846		790	
	(17.035)	(15.992)				
St. dev. students' pre-experiment test scores	0.816	0.832	0.165		770	
	(0.191)	(0.146)				

Note: Means and standard deviations net of randomization block fixed effects. Rows without standard deviations are proportions from binary variables. Sample sizes have been rounded to nearest 10 following NCES restricted data reporting rules. Each experiment collected somewhat different student and teacher characteristics: (a) EET, NROC, and TR; (b) ICL, NROC, and TR; (c) NROC; (d) ICL; (e) EET and TR; and (f) EET.

Table 3—Attrition

	Math			Reading		
	Attrited…			Attrited…		
	before pre-test	after pre-test, but before post-test		before pre-test	after pre-test, but before post-test	
	(1)	(2)	(3)	(4)	(5)	(6)
Student attrition						
Treatment	-0.001	-0.003	-0.004	-0.007	-0.015+	-0.016+
	(0.011)	(0.009)	(0.009)	(0.005)	(0.009)	(0.008)
Pre-experiment test score			-0.041**			-0.025**
			(0.006)			(0.006)
Treatment *			-0.003			-0.000
pre-experiment test score			(0.008)			(0.007)
Number of observations	15,020	12,470	12,470	7,340	6,950	6,950
Control attrition rate	0.176	0.122	0.122	0.057	0.081	0.081
Teacher attrition						
Treatment	0.024	-0.000	-0.028	a	-0.006	-0.099*
	(0.020)	(0.011)	(0.037)		(0.009)	(0.043)
Mean pre-experiment test score			-0.021+			-0.023
			(0.012)			(0.018)
Treatment *			0.020			0.020
mean pre-experiment test score			(0.017)			(0.019)
St. dev. pre-experiment test score			-0.054+			-0.087*
			(0.028)			(0.042)
Treatment *			0.029			0.114*
St. dev. pre-experiment test score			(0.044)			(0.050)
Number of observations	430	400	400	370	370	370
Control attrition rate	0.024	0.015	0.015	0.000	0.012	0.012

Note: Pooled sample of all products and experiments. Each column within a panel reports estimates from a separate linear probability model. In Columns 1 and 4, the dependent variable is an indicator = 1 if the student (teacher) does not have a pre-experiment test score (any students with a pre-experiment test score). In Columns 2-3 and 5-6, the dependent variable is identical except that post-experiment test score replaces pre-experiment. In addition to the dependent variable reported in the table, all regressions include randomization block fixed effects. Sample sizes have been rounded to nearest 10 following NCES restricted data reporting rules.

[a] Zero reading teachers attrited before the pre-test.

+ indicates $p < 0.10$, * 0.05, and ** 0.01

Table 4—Treatment effect on the variance of teacher productivity

	A: Conditional st. dev. of teacher fixed effects			*B:* MLE of teacher st. dev. (random effects)			Number of observations	
	Cont. st. dev.	Treat. - Cont. diff.	Test diff. = 0 p-value	Cont. st. dev.	Treat. - Cont. diff.	Test diff. = 0 p-value	Teachers	Students
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mathematics	0.295	-0.079	0.031	0.315	-0.138	0.000	400	11,030
Reading	0.115	0.002	0.969	0.127	0.014	0.699	360	6,450

Note: Pooled sample of all products and experiments. Teacher productivity measured in student standard deviations of the underlying student tests. Student test scores standardized within tests using control group mean and standard deviation. Column Group A estimated in two steps: (i) estimate within-school shrunken teacher fixed effects controlling for a quadratic in prior test score interacted with indicators for cells formed by combination of subject, grade-level, experiment, and test; (ii) estimate the conditional variance of the estimated teacher fixed effects. The latter step is a least-squares regression of squared residuals on a treatment indicator and randomization block fixed effects; the residuals are obtained from a regression of teacher fixed effects on the same right hand side variables. Standard errors allow for clustering within schools. Column Group B estimated with a linear mixed model including a fixed quadratic in prior test score interacted with cells, school fixed effects, and random effects variance parameters separately for treatment and control teachers. Sample sizes have been rounded to nearest 10 following NCES restricted data reporting rules. As discussed in the text, the study of ICL randomized classes not teachers. In this analysis teacher performance in each class is the unit of analysis for ICL observations, and Column 7 counts 140 classes for ICL.

Table 5—Treatment effects on teachers' use of class time and teachers' in-class effort from classroom observation data

	Math		Reading	
	Control mean	Treat. - Cont. diff	Control mean	Treat. - Cont. diff
	(1)	(2)	(3)	(4)
Use of CAI during class (binary)	0.150	0.644**	0.171	0.789**
		(0.055)		(0.041)
Use of class time				
Proportion of class time spent on…				
Lecturing, whole-class instruction	0.612	-0.296**	0.607	-0.306**
		(0.057)		(0.047)
Individual student work	0.377	0.348**	0.376	0.381**
		(0.053)		(0.039)
Group work	0.092	-0.024	0.214	-0.076*
		(0.040)		(0.037)
Proportion of class time in multiple tasks	0.190	0.019	0.256	0.006
		(0.056)		(0.044)
Teachers' in-class effort				
Proportion of class time the teacher was… (EET)				
Leading	0.599	-0.355**	0.544	-0.202**
		(0.059)		(0.049)
Facilitating	0.261	0.305**	0.369	0.092*
		(0.063)		(0.041)
Monitoring	0.164	0.020	0.256	0.006
		(0.035)		(0.044)
"90+ percent of students on-task" (EET)	0.597	0.205**	0.862	-0.039
		(0.072)		(0.034)
"High level of student attention" (NROC)	0.612	0.205+		
		(0.104)		

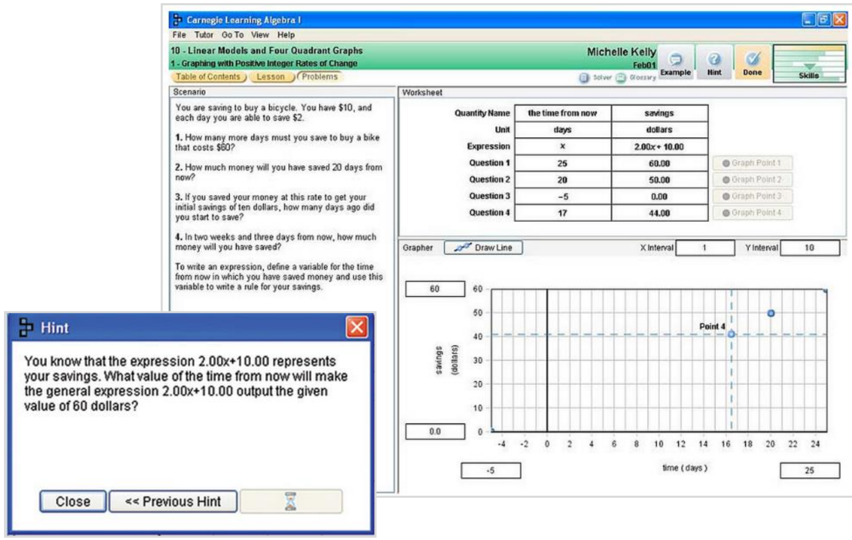
Note: Each cell in Columns 2 and 4 reports a treatment effect (mean) estimate from a separate least-squares regression. Each dependent variable is a proportion or binary indicator. Each regression includes a treatment indicator and randomization block fixed effects. Standard errors allow for clustering within schools. Columns 1 and 3 report control means of the dependent variable net of randomization block fixed effects. Pooled sample of EET, NROC, and TR products and experiments for use of CAI and use of class time estimates, with 230 math teacher observations and 340 reading teacher observations. Estimates for teacher's role (lead, facilitate, monitor) from EET only, with 150 math observations and 270 reading observations. Sample sizes have been rounded to nearest 10 following NCES restricted data reporting procedures. Results separately by study are available in the Appendix Table C2.
+ indicates $p < 0.10$, * 0.05, and ** 0.01

Table 6—Treatment effects on teacher hours worked per week for one class, math teachers

	Control mean hours (st. dev.)	Treat. - Cont. difference log hours (st. err.)	Treat. - Cont. difference in log hours (st. err.) Bottom tercile	Middle tercile	Top tercile
	(1)	(2)	(3)	(4)	(5)
Total work hours	8.997	-0.234**	-0.079	-0.161	-0.525*
	(2.660)	(0.074)	(0.171)	(0.196)	(0.234)
Work hours spent…					
Teaching students	2.961	0.009	0.175	-0.004	-0.184
	(1.149)	(0.084)	(0.161)	(0.303)	(0.211)
Planning lessons	2.617	-0.379**	-0.241	-0.413	-0.587+
	(1.724)	(0.121)	(0.347)	(0.268)	(0.295)
Grading	2.548	-0.316**	-0.159	-0.285	-0.572*
	(1.313)	(0.104)	(0.363)	(0.219)	(0.267)
Testing students	0.871	-0.023	0.386	-0.309	-0.153
	(0.371)	(0.093)	(0.233)	(0.255)	(0.284)

Note: Each cell in Column 2 reports a treatment effect (mean) estimate in log units from a separate least-squares regression. Each dependent variable is log hours worked. Each regression includes a treatment indicator and randomization block fixed effects. Standard errors allow for clustering within schools. Column 1 reports control mean hours worked net of randomization block fixed effects. The three estimates in each row of Columns 3-5 come from a single regression identical to Column 2 except that the treatment indicator is interacted with indicators for the terciles of teacher productivity; terciles defined separately for treatment and control teachers. See text for details of teacher productivity estimates, and identification assumptions for Columns 3-5. The estimation sample for includes 150 math teachers all from the EET study. Sample sizes have been rounded to nearest 10 following NCES restricted data reporting procedures. Standard errors allow for clustering within schools.
+ indicates p < 0.10, * 0.05, and ** 0.01

Panel A—*Cognitive Tutor Algebra I*



Panel B—*Waterford Early Reading*

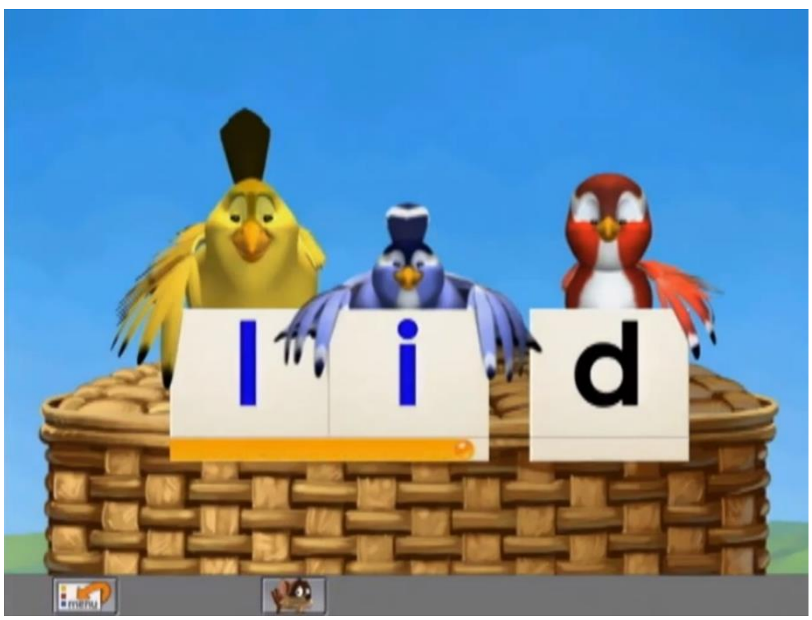
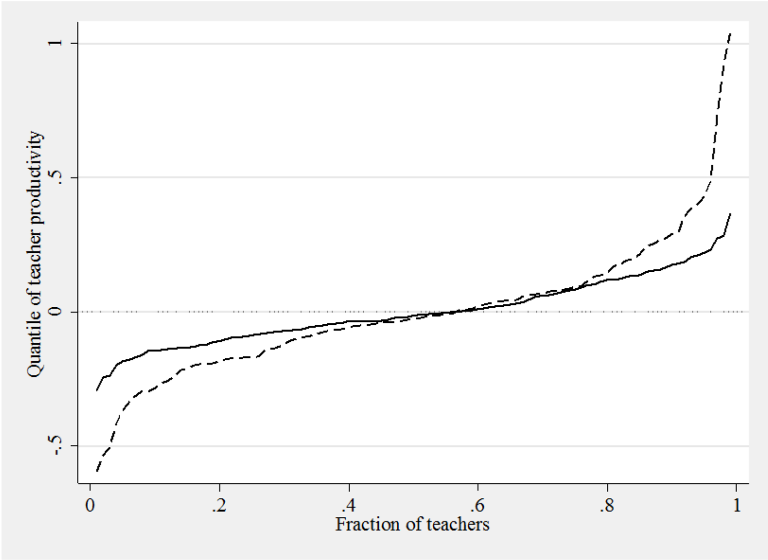


Figure 1—Screen images from CAI software

Note: Panel A shows *Cognitive Tutor Algebra I* © Carnegie Learning, image drawn from and additional examples available at www.carnegielearning.com/galleries/4/ (last accessed November 1, 2013). Panel B shows *Waterford Early Reading* © Waterford Institute, image captured from and full lesson shown at youtu.be/POa5djbx_WY (last accessed November 1, 2013).

Panel A—Math teachers



Panel B—Reading teachers

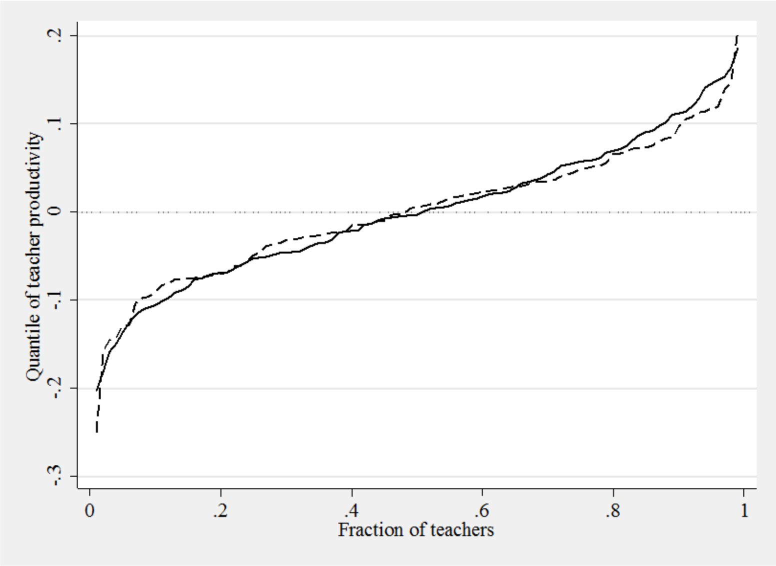


Figure 2—Quantiles of teacher productivity

Note: The solid line traces out the quantiles of estimated teacher productivity (teacher fixed effects estimated as described in Section 3) for treatment teachers, and the dotted line the control quantiles. Productivity measured in student standard deviation units. The estimation sample includes 400 math and 360 reading teachers (rounded to nearest 10).

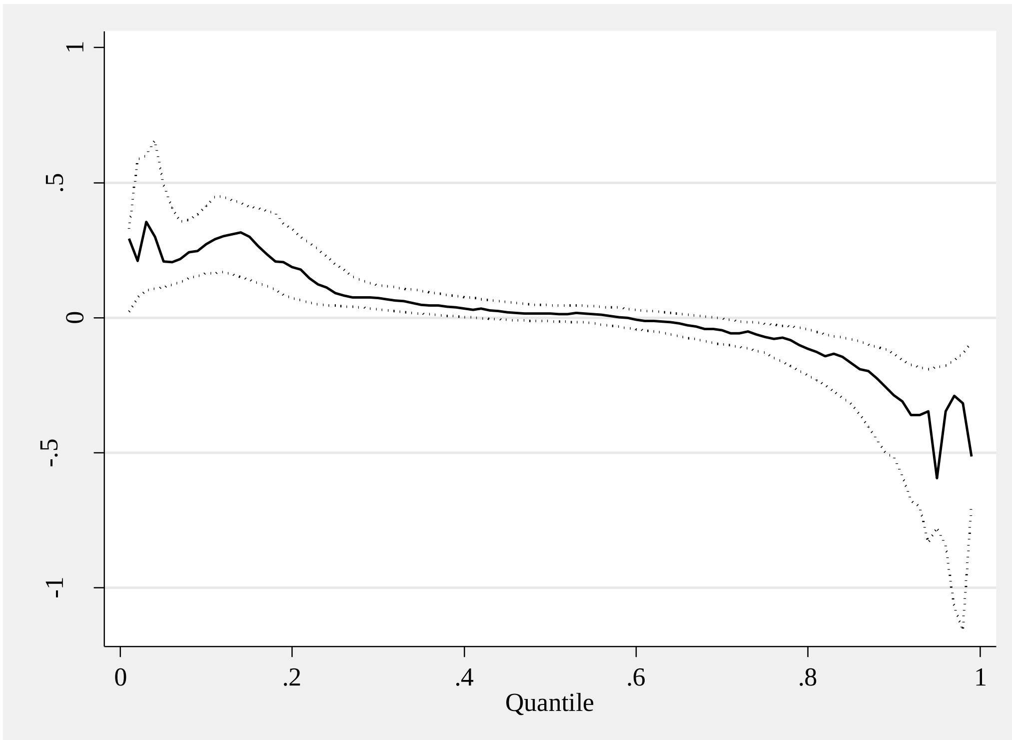


Figure 3—Treatment effects on math teacher productivity by quantile

Note: The solid line traces a series of unconditional quantile treatment effect point estimates, from 0.01 to 0.99 in increments of 0.01. The dotted lines trace the 95 percent confidence intervals. Point estimates are measured in student standard deviation units. Each UQTE is estimated using the recentered influence function approach suggested by Firpo, Fortin, and Lemieux (2009). For each quantile, τ, the dependent variable is the influence function for quantile τ of the distribution of teacher productivity estimates (teacher fixed effects); the independent variables are a treatment indicator (the coefficient plotted) and randomization block fixed effects. Teacher fixed effects estimates come from a regression of standardized post-experiment student test score on a quadratic in pre-experiment test score, school fixed effects, and teacher fixed effects; the parameters of the quadratic are allowed to differ by test. Confidence intervals for the UQTE are cluster-bootstrap standard errors (500 replications) which allow for dependence within schools. The estimation sample includes 400 math teachers (rounded to nearest 10).

NEW TECHNOLOGY AND TEACHER PRODUCTIVITY: APPENDICIES

for online publication

Appendix A: Additional tables

Appendix Table A1—Treatment effects on student test score means

		Re-analysis for this paper			
	Original analysis estimate	no pre-score controls	Obsv.	quadratic in pre-score	Obsv.
	(1)	(2)	(3)	(4)	(5)
Mathematics Pooled		-0.046	12,270	-0.040	11,030
		(0.050)		(0.038)	
EET		0.044	4,540	0.070+	4,060
		(0.073)		(0.038)	
Grade 6	0.07	0.088	3,080	0.125**	2,890
	[0.15]	(0.104)		(0.047)	
Algebra	-0.06	-0.049	1,460	-0.071	1,170
	[0.33]	(0.055)		(0.053)	
ICL (algebra and pre-algebra)	0.173**	0.162*	1,870	0.167**	1,610
	(0.058)	(0.071)		(0.055)	
NROC (algebra)	-0.15	-0.167*	5,860	-0.168*	5,360
	[0.16]	(0.081)		(0.065)	
Reading Pooled		0.000	6,780	0.016	6,450
		(0.033)		(0.017)	
EET		0.027	4,630	0.030	4,300
		(0.039)		(0.021)	
Grade 1	0.03	0.063	2,410	0.051	2,230
	[0.34]	(0.060)		(0.031)	
Grade 4	0.02	-0.012	2,220	0.007	2,070
	[0.48]	(0.047)		(0.027)	
TR (grade 6)	-0.04	-0.056	2,160	-0.011	2,150
	[0.35]	(0.061)		(0.028)	

Note: Standard errors in parentheses, or p-values in brackets when standard errors not reported. Estimates in Column1 taken from original study reports: EET Dynarski et al. (2007), ICL Barrow, Markman, and Rouse (2009), NROC Cavalluzzo et al. (2012), and TR Drummond et al. (2011). Each cell in Columns 2 and 4 reports the treatment effect estimate from a separate student-level least-squares regression using the samples described in the row labels. The dependent variable is standardized post-experiment test score. Column 2 estimates include only randomization block fixed effects, not additional controls. Column 4 estimates include randomization block fixed effects and a quadratic in pre-experiment test score, the parameters of which are allowed to differ by test form. Sample sizes have been rounded to nearest 10 following NCES restricted data reporting procedures. + indicates $p < 0.10$, * 0.05, and ** 0.01

Appendix Table A2—Alternative samples and specifications: treatment effect on the variance of teacher productivity

	A: Conditional st. dev. of teacher fixed effects			*B:* MLE of teacher st. dev. (random effects)			Number of observations	
	Cont. st. dev.	Treat. - Cont. diff.	Test diff. = 0 p-value	Cont. st. dev.	Treat. - Cont. diff.	Test diff. = 0 p-value	Teachers	Students
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mathematics Pooled	0.295	-0.079	0.031	0.315	-0.138	0.000	400	11,030
Excluding EET	0.327	-0.094	0.038	0.350	-0.176	0.001	260	6,970
Excluding ICL	0.231	-0.077	0.088	0.265	-0.124	0.002	260	9,420
Excluding NROC	0.316	-0.067	0.105	0.329	-0.114	0.011	280	5,670
Without school FE[a]	0.330	-0.139	0.040	0.374	-0.221	0.000	400	11,030
Reading Pooled	0.115	0.002	0.969	0.127	0.014	0.699	360	6,450
Excluding EET	0.103	0.008	0.491	0.106	0.033	0.509	90	2,150
Excluding TR	0.116	0.000	0.764	0.135	0.002	0.971	270	4,300
Without school FE[b]	0.115	0.002	0.969	0.127	0.014	0.699	360	6,450

Note: Estimation procedures identical to the main estimates Table 4, except as described in this note. (The main estimates are repeated for convenience in Rows 1 and 6). In Rows 2-4 and 7-8 each row excludes one of the four experiments from the estimation sample. In Rows 5 and 9 the school fixed effects are replaced with randomization block fixed effects, but the sample is the same as the main estimates.

Appendix Table A3—Treatment effects on teacher hours worked per week for one class, reading teachers

	Reading	
	Control	Treat. - Cont. difference
	mean hours (st. dev.)	log hours (st. err.)
	(1)	(2)
Total work hours	4.323	1.319**
	(2.818)	(0.115)
Work hours spent…		
Teaching students	2.309	2.416**
	(2.277)	(0.208)
Planning lessons	1.017	2.306**
	(1.036)	(0.211)
Grading	0.565	2.261**
	(0.819)	(0.197)
Testing students	0.432	1.447**
	(0.511)	(0.189)

Note: Estimation procedure is identical to Columns 1 and 2 in Table 6, except that this table reports results for reading teachers. The sample includes 260 readings teachers all from the EET study. Sample sizes have been rounded to nearest 10 following NCES restricted data reporting procedures. Standard errors allow for clustering within schools.

+ indicates $p < 0.10$, * 0.05, and ** 0.01

Appendix B. Theoretical framework

The job of a classroom teacher involves multiple tasks: lecturing, discipline, one-on-one tutoring, communicating with parents, grading, writing homework problems, and many more. If a computer replaces teachers in performing one of those tasks, then the between-teacher variation in the productivity of that particular task should shrink considerably. Imagine a single pre-recorded video lecture replacing individual teachers' diverse lecture styles and skills. Nevertheless many tasks will remain for each teacher to perform by herself. The total effect of some new technology on the variance of teacher productivity will depend on how individual teachers choose to reallocate time and effort across different tasks after giving some task(s) to the computer.

Consider a teacher deciding how to allocate in-class time across different instructional activities: lectures, group work, individual student work, quizzes. The amount of class time is fixed. A new computer technology which replaces teacher labor and skills in one of those activities, call it activity A, will prompt the teacher to reallocate class time for two notable reasons: First, the computer will change the marginal productivity of class time spent in activity A, compared to what the marginal productivity would be if the teacher used only her own skills. Using the computer may be more effective or less effective than the teacher working alone. Thus the relative productivities of activities A, B, C, etc. will change. Second, the amount of teacher effort required during activity A will likely fall when the computer is replacing teacher labor. Effort is a first-order cost to the teacher of allocating class time to different activities, and thus the relative marginal costs of activities A, B, C, etc. will also change after the new computer technology is available.

To formalize these general observations and to guide my empirical analysis, I propose a version of the teachers' problem that (i) makes a clear distinction between the *tasks* that comprise the job of a classroom teacher, and a teacher's *skills* in each of those tasks; and (ii) explicitly considers the teacher's own decisions about education production in her classroom. The task-skills distinction is a useful and increasingly common feature in the literature on how technical change affects labor (Acemoglu and Autor 2011, 2012). After setting up the basic teacher's problem, I consider the introduction of a computer technology which replaces teacher labor in one task, focusing on how the new technology changes teachers' job decisions and teacher productivity.[1]

A. The teacher's problem and variation in teacher productivity

The responsibility of a classroom teacher is to increase the knowledge and skills of her students. Let m be a measure of teacher productivity: the quantity of knowledge and skill growth attributable to the teacher's work, given the students and other resources she has been allocated by the school.[2] (In this paper's empirical analysis I will measure m using teacher contributions to student math and reading test score growth, but in this theoretical discussion m can be thought of more broadly.) Teacher productivity is a function of both a teacher's skills, $\boldsymbol{\theta}$, and her decisions, $\boldsymbol{x}$, about how to allocate resources—principally her own time and her students' time—across different tasks, $m(\boldsymbol{x}, \boldsymbol{\theta})$, with $\frac{\partial m}{\partial x_k} > 0$ and $\frac{\partial^2 m}{\partial x_k^2} < 0$. Each teacher makes many production decisions, $\boldsymbol{x}' = (x_1, x_2, \dots, x_K)$, including (i) how to allocate a fixed amount of class time across different in-class activities; (ii) what homework to assign students; and (iii) how much time outside of class to spend in other teacher tasks: planning lessons, communicating with parents, grading, extra tutoring, etc. The teacher's skills in each task, $\boldsymbol{\theta}' = (\theta_1, \theta_2, \dots, \theta_K)$, are her stock of current capabilities whether innate, or acquired by training or experience, or both. I incorporate teacher effort shortly.

Variation in teacher productivity, $var(m)$, can arise from differences in teachers' skills, $\boldsymbol{\theta}$, or differences in teachers' input decisions, $\boldsymbol{x}$, or some combination. A large and consistent empirical literature estimates substantial between-teacher variation in m, but, by contrast, very little is known about the functional form of $m(\boldsymbol{x}, \boldsymbol{\theta})$ or the differences in teachers' decisions.

[1] Computer technology could, alternatively, complement or augment teacher skills. In this paper I focus on a technology which is a substitute for teacher labor because that is the nature of the computer-aided instruction applications I study empirically.

[2] In practice, inputs to m like students and other resources vary between teachers in important ways; that variation has to be accounted for when measuring productivity. For this section, consider a representative set of teachers with homogeneous resource allocations from a representative school—including, importantly, homogeneity in the mix of students which they are assigned to teach. The "input decisions" I describe in this section are teachers' decisions about how to use the inputs they have been allocated. I intend for m to represent the teacher's contribution to student outcomes, not the more general education production function to which m is one input. In the empirical analysis I control for differences in students assigned and differences in school-level inputs.

How does a teacher choose $\boldsymbol{x}$? In short, the teacher's preferred decisions, $\boldsymbol{x}^*$, balance a tradeoff between the intrinsic job satisfaction of seeing her students succeed and the disutility of her own work effort.

Consider the teacher's utility maximization problem

$$\max_{x} \; U[w, m(\boldsymbol{x}, \boldsymbol{\theta}), e(\boldsymbol{x}, \boldsymbol{\epsilon})],$$

where U is (i) increasing in her wages, w, and in the achievement and skills she fosters in her students, m; but (ii) decreasing in the effort she must expend, e. Each different input bundle, $\boldsymbol{x}$, requires a different amount of effort from the teacher herself, determined by the function $e(\boldsymbol{x}, \boldsymbol{\epsilon})$. The vector $\boldsymbol{\epsilon}' = (\epsilon_1, \epsilon_2, \dots, \epsilon_K)$ represents the teacher's "effort prices" for each task; prices which vary between teachers.[3] Disutility of effort is common in the analysis of employee behavior across jobs and sectors. By contrast the roles of w and m are peculiar to the teacher's problem. A teacher gains utility from her job performance, but not because performance affects compensation or continued employment as it would in other jobs. In practice, teacher compensation, w, does not depend on skills, job decisions, or productivity; and involuntary turnover is rare.[4] Teachers are, however, generally considered "motivated agents" who derive utility directly from contributing to their students' growth and success, m, (Dixit 2002).

I focus here on teachers input decisions, $\boldsymbol{x}$, holding other things fixed. This is a plausible description of the teacher's problem in the "short run"—for example, the duration of one school year. I assume that, in the short run, the teacher takes as fixed her own skills, $\boldsymbol{\theta}$; her own costs, $e(\cdot)$; her wages, w; and her job assignment (i.e., the subject, grade, and specific group(s) of students she is assigned to teach). Teachers may be able to change their skills, job assignments, and other variables in the long run, but decisions like how much time to spend lecturing or how much homework to assign are made day to day and week to week.

This framework suggests individual teachers' production decisions, and the tradeoff between m and e those decisions make, are important considerations when studying teacher productivity in general. But examination of teachers' production decisions by economists has

[3] In this version of the teacher's problem effort is not an explicit or separate choice variable for the teacher. Rather the teacher makes her choice $\boldsymbol{x}$ knowing that $\boldsymbol{x}$ will require effort determined by the function e. The function e may differ from teacher to teacher.

[4] Some recent, publically-noted teacher contracts explicitly make low performance a determinant of dismissal, or make wages a function of performance. The utility function in 1 could be modified replacing w with an expected wage, e.g. $w\mathbb{E}[m,\cdot]$, or more general relationship, e.g. $w(m,\cdot)$. The general result—that teachers trade off student achievement and cost of effort—would still hold. I stay with the traditional fixed wages for simplicity of exposition.

been rare (Murnane and Phillips 1981; Brown and Saks 1987; and Betts and Shkolnik 1999 are exceptions).

B. The teacher's response to new technology

In this paper I focus on how $\boldsymbol{x}$ and m change when a new labor-replacing computer technology is made available to teachers. To make the discussion concrete, consider a simple example of the teacher's problem where the teacher makes one decision: how to allocate a fixed amount of class time, $\bar{t}$, between two activities x_1 and x_2, for example, individual student work and lecturing. Let m have a constant elasticity of substitution form, e a simple linear form, and let w, m and e be linearly separable. The teacher's problem before computers is

$$\max_{x_1,x_2} \quad ([\theta_1 x_1]^\rho + [\theta_2 x_2]^\rho)^{\frac{1}{\rho}} - [\epsilon_1 x_1 + \epsilon_2 x_2] \qquad \text{subject to } x_1 + x_2 \leq \bar{t}. \tag{B1}$$

The teacher chooses (x_1^*, x_2^*), uses her own skills (θ_1, θ_2) to carry out the work, pays the effort costs e^*, and produces m^*.

Now imagine a new computer tool which can replace teacher labor in performing task x_1. The computer has some level of "skill", ϕ_1, in performing that task. When the new technology is available for the teacher to use, the teacher's problem becomes

$$\max_{x_1,x_2,c} \quad \left(\left[\tilde{f}(\theta_1, c)x_1\right]^\rho + [\theta_2 x_2]^\rho\right)^{\frac{1}{\rho}} - [\tilde{g}(\epsilon_1, c)x_1 + \epsilon_2 x_2], \quad \text{subject to } x_1 + x_2 \leq \bar{t}$$
$$\text{where } \tilde{f}(\theta_1, c) = c\phi_1 + (1-c)\theta_1, \ \ c \in \{0,1\}. \tag{B2}$$

Now, in addition to choosing the quantity of x_1 and x_2, the teacher must also choose whether to use the computer, $c = 1$, or not, $c = 0$.[5] Let $\widetilde{m}^*$ be teacher productivity with the new computer tool available, and similarly other variables with the $\sim$ notation.

After the introduction of the new technology the variance of teacher productivity will shrink, $var(\widetilde{m}^*) \leq var(m^*)$, if the computer is a substitute for teacher skills and effort. Stated formally,

[5] I assume that the availability of new technology does not change the set of tasks the school assigns to the teacher. The teacher remains responsible for x_1 whether the computer is used or not. To date, there are only very limited examples of schools redesigning teachers' jobs tasks as a result of new technology.

PROPOSITION 1. *The introduction of a new computer technology will reduce the variance of teacher productivity,* $var(\tilde{m}^*) \leq var(m^*)$, *if*

(i) *the computer is a substitute for teacher skills in some task,* $0 \leq \frac{\partial \tilde{f}}{\partial \theta_1} \leq 1$, *and*

(ii) *the computer is a substitute for teacher effort in the same task,* $0 \leq \frac{\partial \tilde{g}}{\partial \epsilon_1} \leq 1$,

(iii) *but the computer does not otherwise change the underlying production function or effort cost function, that is if* $m = n(\boldsymbol{x}, \theta_1, \theta_2, \dots, \theta_K)$ *then* $\tilde{m} = n\left(\boldsymbol{x}, \tilde{f}(\theta_1), \theta_2, \dots, \theta_K\right)$ *and similarly for e and ẽ.*

A proof is shown below.[6] Intuitively, before the computer, differences between teachers in the productivity of time spent on task x_1 were due to differences in teachers' skills θ_1. The computer option weakens the relationship between skill and productivity, $\tilde{f}' \leq 1$, and thus shrinks the differences in the productivity of time spent on task x_1. Additionally, as I discuss shortly, the computer option will encourage (many) teachers to reallocate more class time to x_1 affecting the magnitude of the reduction in variance. The reallocation is furthered by the changes in effort costs, $\tilde{g}' \leq 1$.

The conditions of Proposition 1 certainly do not describe all educational computer technologies, but those conditions do (plausibly) describe computer-aided instruction software. CAI is designed to take on the teacher's role in one-on-one teacher-student tutoring—substituting the computer's skills for the teacher's. Yet, while CAI might make some educational activities more or less productive, the structure of the educational production function is not radically changed. By contrast, the conditions of Proposition 1 would not hold for technology which complements teacher skills, $\tilde{f}' > 1$, for example an interactive white board.

The total effect of new technology on teacher productivity can be separated, at least conceptually, into a *standardization effect* and a *reallocation effect.* Standardization is the more intuitive effect: machines standardize the execution of the tasks they are given to perform, while human performance of the same tasks would inherently vary. The standardization effect would be the only effect if the introduction of new technology did not change the teacher's time allocation decisions, that is if $(x_1^*, x_2^*) = (\tilde{x}_1^*, \tilde{x}_2^*)$ in the current example. However, a rational

[6] The prediction and proof apply to a range of production and cost functions, including but not limited to the CES-style m. The prediction and proof can also easily be extended to a technology which replaces labor in several tasks. Finally, the specific $\tilde{f}$ in Problem 2 meets condition (i) but, of course, other $\tilde{f}$ would as well.

teacher will reevaluate her preferred time allocation decisions, and the resulting reallocation will contribute to what we observe as the total change in productivity.

New technology will prompt (many) teachers to reallocate time for two reasons: changes in productivity and changes in costs. First, using the computer for task x_1 will change the marginal rate of technical substitution of x_1 for x_2. The teacher should, all else equal, shift class time away from x_2 and into x_1 if the computer is more productive than she is herself. Using the maximizing solutions to problems 1 and 2,

$$\tilde{x}_1^* > x_1^* \quad \Leftrightarrow \quad \left(\frac{\tilde{f}(\theta_1,c)}{\theta_1}\right)^{\rho} > \frac{\tilde{g}(\epsilon_1,c)}{\epsilon_1}. \tag{B3}$$

Thus, holding effort costs constant, for teachers with $\phi_1 > \theta_1$ we would expect that $\tilde{c}^* = 1$, $\tilde{x}_1^* > x_1^*$, and $\tilde{x}_2^* < x_2^*$. By contrast, teachers who are more skilled than the computer, $\phi_1 < \theta_1$, could choose to ignore the new tool leaving $(x_1^*, x_2^*) = (\tilde{x}_1^*, \tilde{x}_2^*)$. In short, on this productivity margin, we would predict an increase in average time allocated to x_1, complementing the standardization effect, and thus increasing the magnitude of the reduction in the variance of teacher productivity.

Second, teachers will also be prompted to reallocate time if using the computer changes the marginal costs of x_1. The primary marginal cost to the teacher is the effort required of her, e, to carry out $\boldsymbol{x}$. The teacher should, all else equal, shift class time away from x_2 and into x_1 if using the computer reduces the effort she must expend: $\tilde{g}(\epsilon_1, c) < \epsilon_1$ in Equation B3.

The net change in x_1 for any one teacher, then, is a race between changes at these two margins: productivity and costs. Whether x_1 will increase on average, $\mathbb{E}[\tilde{x}_1^*] > \mathbb{E}[x_1^*]$, will depend the baseline distribution of skills and effort costs among teachers, and the particular computer tool. One intuitive case where $\mathbb{E}[\tilde{x}_1^*] > \mathbb{E}[x_1^*]$ is described in Proposition 2.

PROPOSITION 2. *The introduction of a new computer technology will increase the average amount of class time allocated to task* x_1, $\mathbb{E}[\tilde{x}_1^*] > \mathbb{E}[x_1^*]$, *if* U *is as specified in problem 2 and using the computer*

(i) *increases the marginal productivity of task* x_1 *on average,* $\mathbb{E}\left[\frac{\tilde{f}(\theta_1,1)}{\theta_1}\right] > 1$, *and*

(ii) *reduces the marginal costs of task* x_1 *on average,* $\mathbb{E}\left[\frac{\tilde{g}(\epsilon_1,1)}{\epsilon_1}\right] < 1$

among teachers who choose to use the computer, $c^* = 1$.

A proof is shown below.[7]

Conditions (i) and (ii) in Proposition 2 are a plausible description of computer-aided instruction software. First, a lab of computers can tutor multiple students at once, giving each student full attention. A teacher, by contrast, must divide her tutoring attention across multiple students (Barrow, Markman, and Rouse 2008). Thus the computer could deliver a higher "dose" of tutoring even if the teacher is more skilled at tutoring in absolute terms. Second, the teacher may need to be present for the same amount of class time, $\bar{t}$, but with less demanding responsibilities during class or less planning required before class. Of course the reverse could be true; using CAI might increase the effort required for x_1, and, all else equal, cause the teacher to increase x_2. For example, students using computers might make maintaining discipline more work for the teacher. Finally, conditions (i) and (ii) need only hold among teachers who adopt CAI, $c^* = 1$. The teacher for whom CAI reduces the marginal productivity of x_1 and increases its marginal costs could (should) simply ignore the new tool leaving her behavior unchanged, $\tilde{x}_1^* = x_1^*$. In short, the introduction of CAI should increase the average amount of class time teachers allocate to students working individually. In the empirical analysis I test for changes in class time allocation, and changes in the marginal effort costs of teachers.

A third proposition describes a simple but notable observation in this framework; it concerns teachers whose productivity would fall if they use the computer tool.

PROPOSITION 3. *A rational teacher will choose to adopt the new technology, $c^* = 1$, even if her productivity falls, $\widetilde{m}^* < m^*$, if the reduction in her productivity is smaller, in utility terms, than the reduction her in effort costs.*[8]

This observation is simple but it underscores the importance of considering both m and e in deciding how to optimally manage the teacher workforce. Some students may be worse off—they learn less at school—if their teacher adopts a new technology even while their teacher is made better off.

[7] Proposition 2 is a specific case of a more general proposition that $\mathbb{E}[\tilde{x}_1^*] > \mathbb{E}[x_1^*]$ if U is as specified in problem 2 and $\mathbb{E}\left[\frac{\tilde{f}(\theta_1,1)}{\theta_1}\right] > \frac{1}{\rho}\mathbb{E}\left[\frac{\tilde{g}(\epsilon_1,1)}{\epsilon_1}\right] | c^* = 1$.

[8] The proof is straightforward. Assume that w, m, and e are linearly separable in U. In problem 2,
$c^* = 1 \iff (\widetilde{m}^* + \tilde{e}^* | c = 1) \geq (\widetilde{m}^* + \tilde{e}^* | c = 0) = (m^* + e^*) \iff (\widetilde{m}^* - m^*) \geq (\tilde{e}^* - e^*)$.
The last inequality could, of course, hold even if $(\widetilde{m}^* - m^*) < 0$.

C. Additional considerations

Two additional considerations are worth discussing in this section. First, to solve problems like 2, teachers need good information about their own skills and the computer's "skills", among other things. Overly pessimistic beliefs about herself or optimistic beliefs about the technology would lead the teacher to overuse the technology. For example, in this paper's empirical setting, any positive recommendation of CAI implied by the experimental trial itself could lead teachers to overestimate the value of CAI for their classroom. If teachers overuse (underuse) the new technology, the reductions in the variance of teacher productivity would be larger (smaller) and the increases in x_1 similarly larger (smaller).

Second, teachers may feel pressure to use the new technology even if doing so would reduce their own utility or productivity; pressure from their managers, other teachers, or researchers. The resulting overuse would make the reductions in the variance of teacher productivity larger than would be the case if teachers freely chose. The effect on average time allocated to x_1 is less clear, depending importantly on whether the pressure extends to class time allocation decisions or not.[9]

D. Proofs

Proposition 1 is a special case of Lemma 1.

Lemma 1. Let $\mathbf{Z}' = (\mathbf{\Theta}', \mathbf{E}') = (\Theta_1, \Theta_2, \ldots, \Theta_K; E_1, E_2, \ldots, E_K)$ be a vector of $J = 2K$ exogenous random variables, teacher skills and teacher effort prices in the current paper, where each $z_j \in [a_j, b_j]$. Let $m^*: \mathbf{z} \to \mathbb{R}^+$ and $\widetilde{m}^*: \mathbf{z} \to \mathbb{R}^+$ be differentiable functions. For notation, let h_i' be the partial derivative of a function h with respect to its i-th argument, e.g., $m_1^{*\prime} \equiv \frac{\partial m}{\partial \theta_1}$. If

(i) $m^*(\mathbf{z}) = n(f(\theta_1), \theta_2, \ldots, \theta_K, g(\epsilon_1), \epsilon_2, \ldots, \epsilon_K)$ and
$\widetilde{m}^*(\mathbf{z}) = n\left(\tilde{f}(\theta_1), \theta_2, \ldots, \theta_K, \tilde{g}(\epsilon_1), \epsilon_2, \ldots, \epsilon_K\right)$,

(ii) $0 \le \tilde{f}' \le f'$, and

(iii) $0 \le \tilde{g}' \le g'$

[9] A third consideration: The discussion to this point has focused on teachers' decisions in one school year, indeed the first year the technology is available, but teachers may be maximizing a stream of expected outcomes across the current and future years. A rational teacher may, therefore, begin using technology even if her productivity will fall in year one, if she believes the short-run costs are an investment in future, offsetting productivity gains. The empirical data for this study are limited to a single school year. However, in a multi-year model the teacher should inter-temporally smooth utility, to some extent, by increasing effort in year one; the available data do allow me to test that prediction.

then

$$var\big(\widetilde{m}^*(\mathbf{Z})\big) \leq var\big(m^*(\mathbf{Z})\big).$$

Proof.

First, condition (a) of Lemma 2 holds. By assumption (i) $m_1^{*\prime} = n_1' * f'$ and $\widetilde{m}_1^{*\prime} = n_1' * \tilde{f}'$, noting that $\frac{\partial z_i}{\partial z_j} = 0$ for all $j \neq i$. By assumption (ii) $sgn\big(\tilde{f}'\big) = sgn(f')$. Thus $sgn(\widetilde{m}_1^{*\prime}) = sgn(m_1^{*\prime})$. By a similar argument, using assumption (iii), $sgn(\widetilde{m}_{K+1}^{*\prime}) = sgn(m_{K+1}^{*\prime})$. Finally, $m_j^{*\prime} = \widetilde{m}_j^{*\prime} \Rightarrow sgn\big(\widetilde{m}_j^{*\prime}\big) = sgn\big(m_j^{*\prime}\big)$ for all $j \neq 1, (K+1)$.

Second, condition (b) of Lemma 2 holds. Assumption (ii), $0 \leq \tilde{f}' \leq f', \Longrightarrow$

$$\big|\tilde{f}'\big| \leq |f'| \Longrightarrow |n_1'| * \big|\tilde{f}'\big| \leq |n_1'| * |f'| \Longrightarrow |\widetilde{m}_1^{*\prime}| \leq |m_1^{*\prime}|.$$

By a similar argument, using assumption (iii), $|\widetilde{m}_{K+1}^{*\prime}| \leq |m_{K+1}^{*\prime}|$. Finally, again, $m_j^{*\prime} = \widetilde{m}_j^{*\prime} \Rightarrow \big|\widetilde{m}_j^{*\prime}\big| = \big|m_j^{*\prime}\big|$, for all $j \neq 1, (K+1)$.

Applying Lemma 2 completes the proof.

□

Lemma 2. Let $\mathbf{Z}' = \big(Z_1, Z_2, \dots, Z_J\big)$ be a vector of J random variables, where each $z_j \in \big[a_j, b_j\big]$. Let $f: \boldsymbol{z} \to \mathbb{R}^+$ and $g: \boldsymbol{z} \to \mathbb{R}^+$ be differentiable functions, and $f_j' \equiv \frac{\partial f}{\partial z_j}$. If for all j

(a) $sgn\big(f_j'\big) = sgn\big(g_j'\big)$, and

(b) $\big|f_j'\big| \leq \big|g_j'\big|$, then

$$var\big(f(\mathbf{Z})\big) \leq var\big(g(\mathbf{Z})\big).$$

Proof.

Define a new function $h(\boldsymbol{z}, \lambda) = \lambda f(\boldsymbol{z}) + (1-\lambda) g(\boldsymbol{z})$, where $\lambda \in [0,1]$, then

$$\frac{\partial}{\partial \lambda} var\big(h(\mathbf{Z}, \lambda)\big) \leq 0 \Rightarrow var\big(h(\mathbf{Z}, 1)\big) \leq var\big(h(\mathbf{Z}, 0)\big) \Rightarrow var\big(f(\mathbf{Z})\big) \leq var\big(g(\mathbf{Z})\big).$$

$$\frac{\partial}{\partial \lambda} var\big(h(\mathbf{Z}, \lambda)\big)$$

$$= \frac{\partial}{\partial \lambda}\{\mathbb{E}[h(\mathbf{Z}, \lambda)^2] - \mathbb{E}[h(\mathbf{Z}, \lambda)]^2\}$$

apply Leibnitz-Reynolds' Theorem, noting that a_j and b_j are constants $\forall\ j$

$$= \mathbb{E}\left[\frac{\partial}{\partial \lambda} h(\mathbf{Z}, \lambda)^2\right] - 2\mathbb{E}[h(\mathbf{Z}, \lambda)]\mathbb{E}\left[\frac{\partial}{\partial \lambda} h(\mathbf{Z}, \lambda)\right]$$

$$= 2\mathbb{E}[h(\mathbf{Z}, \lambda) h'(\mathbf{Z}, \lambda)] - 2\mathbb{E}[h(\mathbf{Z}, \lambda)]\mathbb{E}[h'(\mathbf{Z}, \lambda)]$$

notice that, for any given $\lambda \in [0,1]$,

(a) if f_j' and $g_j' \geq 0$, then h is an increasing function of z_j, and $h' \equiv \frac{\partial}{\partial \lambda} h(\mathbf{z}, \lambda)$ is a decreasing function of z_j:

$$\frac{\partial}{\partial z_j} h(\mathbf{z}, \lambda) = \lambda f_j' + (1-\lambda) g_j' \geq 0$$

$$\frac{\partial}{\partial z_j}\left[\frac{\partial}{\partial \lambda} h(\mathbf{z}, \lambda)\right] = \frac{\partial}{\partial z_j}[f - g] = f_j' - g_j' \leq 0$$

(b) else if f_j' and $g_j' \leq 0$, then h is a decreasing function of z_j, and $h'$is an increasing function of z_j;

by Chebyshev's Order Inequality $\mathbb{E}[h(\mathbf{Z}, \lambda) h'(\mathbf{Z}, \lambda)] \leq \mathbb{E}[h(\mathbf{Z}, \lambda)]\mathbb{E}[h'(\mathbf{Z}, \lambda)]$, thus

≤ 0

□[10]

Proposition 2. The introduction of a new computer technology will increase the average amount of class time allocated to task x_1, $\mathbb{E}[\tilde{x}_1^*] > \mathbb{E}[x_1^*]$, if U *is* as specified in problem 2 and using the computer

(i) increases the marginal productivity of task x_1 on average, $\mathbb{E}\left[\frac{\tilde{f}(\theta_1, 1)}{\theta_1}\right] > 1$, and

(ii) reduces the marginal costs of task x_1 on average, $\mathbb{E}\left[\frac{\tilde{g}(\epsilon_1, 1)}{\epsilon_1}\right] < 1$

among teachers who choose to use the computer, $c^* = 1$.

Proof.

$\mathbb{E}[\tilde{x}_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon}) - x_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon})]$

if $c \in \{0,1\}$

$= \mathbb{E}[c^*]\mathbb{E}[\tilde{x}_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon}) - x_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon}) | c^* = 1] + (1 - \mathbb{E}[c^*])\mathbb{E}[\tilde{x}_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon}) - x_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon}) | c^* = 0]$

$= \mathbb{E}[c^*]\mathbb{E}[\tilde{x}_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon} | c = 1) - \tilde{x}_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon} | c = 0) | c^* = 1]$

$= \mathbb{E}[c^*]\mathbb{E}\left[\frac{\partial \tilde{x}_1^*(\boldsymbol{\theta}, \boldsymbol{\epsilon})}{\partial c} | c^* = 1\right]$

$= \mathbb{E}[c^*]\mathbb{E}\left[\frac{\partial x_1^*}{\partial \tilde{f}(\theta_1)} \frac{\partial \tilde{f}(\theta_1)}{\partial c} + \frac{\partial x_1^*}{\partial \tilde{g}(\epsilon_1)} \frac{\partial \tilde{g}(\epsilon_1)}{\partial \epsilon_1} \ | c^* = 1\right]$

[10] A proof by See and Chen (2008) suggested the use of $\frac{\partial}{\partial \lambda} var(h)$. This proof requires weaker assumptions, and allows $\mathbf{Z}$ to be multivariate.

See, C. & Chen, J. (2008). Inequalities on the variances of convex functions of random variables. *Journal in Pure and Applied Mathematics, 9*(3).

from here use the CES production and linear cost functions as specified in problem 2

$$= \mathbb{E}[c^*]\mathbb{E}\left[\{x_1^*(1+x_1^*)\}\left\{\left(\frac{\rho}{1-\rho}\right)\frac{1}{\theta_1}\left(\tilde{f}(\theta_1,1)-\theta_1\right)+\left(\frac{-1}{1-\rho}\right)\frac{1}{\epsilon_1}(\tilde{g}(\epsilon_1,1)-\epsilon_1)\right\}|c^*=1\right]$$

$$= \mathbb{E}[c^*]\mathbb{E}_{x_1^*}\left[\mathbb{E}\left[\{x_1^*(1+x_1^*)\}\left\{\left(\frac{\rho}{1-\rho}\right)\frac{1}{\theta_1}\left(\tilde{f}(\theta_1,1)-\theta_1\right)+\left(\frac{-1}{1-\rho}\right)\frac{1}{\epsilon_1}(\tilde{g}(\epsilon_1,1)-\epsilon_1)\right\}|x_1^*\right]|c^*=1\right]$$

$$= \mathbb{E}[c^*]\mathbb{E}[x_1^*(1+x_1^*)|c^*=1]\mathbb{E}\left[\left(\frac{\rho}{1-\rho}\right)\frac{1}{\theta_1}\left(\tilde{f}(\theta_1,1)-\theta_1\right)+\left(\frac{-1}{1-\rho}\right)\frac{1}{\epsilon_1}(\tilde{g}(\epsilon_1,1)-\epsilon_1)|c^*=1\right]$$

The first two expectation terms in this expression will be positive, thus the sign of the third expectation will determine in the sign of $\mathbb{E}[\tilde{x}_1^*(\boldsymbol{\theta},\boldsymbol{\epsilon})-x_1^*(\boldsymbol{\theta},\boldsymbol{\epsilon})]$. When will the third term be positive?

$$\mathbb{E}\left[\left(\frac{\rho}{1-\rho}\right)\frac{1}{\theta_1}\left(\tilde{f}(\theta_1,1)-\theta_1\right)+\left(\frac{-1}{1-\rho}\right)\frac{1}{\epsilon_1}(\tilde{g}(\epsilon_1,1)-\epsilon_1)|c^*=1\right]>0$$

$$\mathbb{E}\left[\left(\frac{\rho}{1-\rho}\right)\frac{1}{\theta_1}\left(\tilde{f}(\theta_1,1)-\theta_1\right)\right]>-\mathbb{E}\left[\left(\frac{-1}{1-\rho}\right)\frac{1}{\epsilon_1}(\tilde{g}(\epsilon_1,1)-\epsilon_1)\right]|c^*=1$$

$$\mathbb{E}\left[\frac{1}{\theta_1}\left(\tilde{f}(\theta_1,1)-\theta_1\right)\right]>\frac{1}{\rho}\mathbb{E}\left[\frac{1}{\epsilon_1}(\tilde{g}(\epsilon_1,1)-\epsilon_1)\right]|c^*=1$$

$$\mathbb{E}\left[\frac{\tilde{f}(\theta_1,1)}{\theta_1}\right]>\frac{1}{\rho}\mathbb{E}\left[\frac{\tilde{g}(\epsilon_1,1)}{\epsilon_1}\right]|c^*=1$$

□

Appendix C. Combining classroom observation data

In three of four experiments—EET, NROC, and TR—researchers observed teachers and students during class time, and recorded, among other data, what instructional activities took place during class. This appendix describes differences across studies in the observation protocols and data collection instruments, and my decisions in combining the data. Complete details of instrument and protocol development, and other data collected during observations, are available in the original study reports.

Researchers for the EET study, which contributes three-quarters of the classroom observation data in Table 5, observed classes for one hour. The observers' data collection form listed several instructional activities, for example, "Lecture: teacher talking or presenting materials and students are mostly listening" and "Solo independent practice (at desk or computer): i.e., reading silently, worksheets, exercises." Every ten minutes the researcher marked the activity currently in use (or activities if more than one). First, as detailed in Appendix Table C1, I combine the raw data into three task categories: (i) lecturing or whole-class activities, (ii) students working individually, and (iii) students working in pairs or small groups. Second, for each teacher, I calculate the mean of her binary 10-minute-interval data to estimate the proportion of class time in each of the three tasks.[1]

TR and NROC protocols were similar. TR and NROC researchers also observed classes for one hour, and also recorded the current activity (activities) at regular intervals: every seven minutes for TR and every fifteen minutes for NROC. As shown in Appendix Table C1, the TR and NROC data collection form measured the same three main task categories, but using different language and item structure. The TR observation data, like the EET data, include each 7-minute-interval record; and, as with the EET data, I calculate the mean of the interval data. However, the NROC 15-minute-interval records were apparently not kept for analysis. Instead the NROC observation data include only a single categorical measure of frequency for each task, from 0 for "not observed" to 5 for "extensively". For each teacher in the NROC study, I use a binary measure $= 1$ if the task was observed at a frequency of four or five.

[1] EET classroom observers also marked, at each 10 minute interval, the "teacher's role during this activity" from among "leader," "facilitator," or "monitor/observer"; and the "proportion of students not doing the assigned task/activity." Using these data I similarly calculate the proportion of class time the teacher spends in each role, and the proportion of class time with 90 percent or more of students on task.

The results in Table 5 are estimated using the combined data of the EET, NROC, and TR studies. The combined data include one observation per teacher, with a frequency measure for each of the three tasks, created as described in the previous two paragraphs. These frequency measures, y_j in Equation 5 notation, are imperfect measures of a teacher's true class time allocation, y_j^* thus $y_j = y_j^* + \xi_j$. Moreover, the nature of ξ_j is likely to differ from study to study, given the differences in observation protocols and data collection instruments. Nevertheless, the treatment effects in Table 5 are appropriately interpreted as "proportion of class time" under the assumption that $\mathbb{E}[\xi_j | T_j] = \mathbb{E}[\xi_j]$. This assumption would be satisfied, for example, if the times selected for classroom observations were not a function of treatment condition. Additionally, estimates in Table 5, using Specification 5, use only within study variation, and the reported standard errors are heteroskedasticity-cluster robust.

Finally, while this appendix details the specific decisions underlying Table 5, the pattern of time allocation and treatment effects in Table 5 is robust to alternative decisions for processing the raw observation data. For example, turning the EET and TR data into binary measures to match the NROC structure, and alternative cut-offs for making the NROC raw categorical data into binary data. Additionally, as shown in Appendix Table C2, the pattern of results in Table 5 also holds for each of the three studies when analyzed individually.

Appendix Table C1—Combining classroom observation instruments into key task categories

Task	Study	Items included in task category, original instrument text
Lecturing, whole-class instruction	EET	▪ Lecture: teacher talking or presenting material and students are mostly listening ▪ Question and answer: teacher is leader and interaction with students is focused on questioning
	TR	▪ Was the instructional grouping a whole class or large group?
	NROC	▪ Direct instruction (lecture)
Individual student work	EET	▪ Solo independent practice (at desk or computer): i.e., reading silently, worksheets, exercises
	TR	▪ Were students working individually? ▪ Were students working individually with teachers?
	NROC	▪ Independent seatwork (self-paced worksheets, individual assignments) ▪ Computer for instructional delivery (computer-assisted instruction, drill and practice)
Group work	EET	▪ Pair or group practice, problem solving, or project work
	TR	▪ Was the instruction grouping a small group? ▪ Was the instructional grouping in pairs?
	NROC	▪ Cooperative/collaborative learning ▪ Student discussion

Note: The assignment of items to the three categories was informed by reviewing original data collection instruments and observer training materials when available. For additional details on instruments and training see EET Dynarski et al. (2007), TR Drummond et al. (2011), and NROC Cavalluzzo et al. (2012).

Appendix Table C2—Treatment effects on the use of class time, and teachers' in-class effort by study sample

	Math				Reading			
	EET		NROC		EET		TR	
	Control mean	Treat. - Cont. diff	Control mean	Treat. - Cont. diff	Control mean	Treat. - Cont. diff	Control mean	Treat. - Cont. diff
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Use of CAI during class (binary)	0.076	0.808**	0.245	0.366**	0.209	0.771**	0.013	0.859**
		(0.050)		(0.103)		(0.048)		(0.050)
Use of class time								
Proportion of class time spent on…								
Lecturing, whole-class instruction	0.470	-0.288**	0.863	-0.310*	0.563	-0.251**	0.779	-0.530**
		(0.044)		(0.123)		(0.050)		(0.114)
Individual student work	0.307	0.412**	0.487	0.239*	0.380	0.361**	0.368	0.464**
		(0.063)		(0.094)		(0.043)		(0.093)
Group work	0.093	-0.032	0.090	-0.011	0.173	-0.030	0.376	-0.267*
		(0.050)		(0.066)		(0.037)		(0.112)
Proportion of class time in multiple tasks	0.035	0.031	0.464	-0.002	0.198	0.083+	0.479	-0.312**
		(0.033)		(0.126)		(0.045)		(0.106)

Note: Accompanies Table 5. Each cell in even numbered columns reports a treatment effect (mean) estimate from a separate least-squares regression. Each dependent variable is a proportion or binary indicator. Each regression includes a treatment indicator and randomization block fixed effects. Standard errors allow for clustering within schools. Odd numbered columns report control means of the dependent variable net of randomization block fixed effects. EET math sample includes 150 teacher observations. Similarly, NROC math 80, EET reading 270, TR reading 60.Sample sizes have been rounded to nearest 10 following NCES restricted data reporting procedures.
+ indicates p < 0.10, * 0.05, and ** 0.01