Upgrading Education with Technology: Insights from Experimental Research†

MAYA ESCUETA, ANDRE JOSHUA NICKOW, PHILIP OREOPOULOS, AND VINCENT QUAN *

In recent years, there has been widespread interest around the potential for technology to transform learning. As investment in education technology continues to grow, students, parents, and teachers face a seemingly endless array of education technologies from which to choose—from digital personalized learning platforms to online courses to text message reminders to submit financial aid forms. Amid the excitement, it is important to step back and understand how technology can help—or in some cases hinder—learning. This review article synthesizes and discusses rigorous evidence on the effectiveness of technology-based approaches to education in developed countries and outlines areas for future inquiry. In particular, we examine randomized controlled trials and regression discontinuity studies across the following categories of education technology: (i) access to technology, (ii) computer-assisted learning, (iii) technology-enabled behavioral interventions in education, and (iv) online learning. We hope this synthesis will advance academic understanding of how technology can improve education, outline key areas for new experimental research, and help drive improvements to the policies, programs, and structures that contribute to successful teaching and learning. (JEL H52, H75, I20, O33)

1. Introduction

Technological innovation over the past two decades has indelibly altered today’s education landscape. Revolutionary advances in information and communications technology (ICT)—particularly associated with computers, mobile phones, and the internet

* Escueta: Columbia University. Nickow: Northwestern University. Oreopoulos: University of Toronto and NBER. Quan: Abdul Latif Jameel Poverty Action Lab (J-PAL). We are extremely grateful to Caitlin Anzelone, Rekha Balu, Peter Bergman, Brad Bernatek, Ben Castleman, Angela Duckworth, Jonathan Guryan, Alex Haslam, Andrew Ho, Ben Jones, Matthew Kraft, Kory Kroft, David Laibson, Susanna Loeb, Andrew Magliozi, Ignacio Martinez, Susan Mayer, Steve Mintz, Piotr Mitros, Lindsay Page, John Pane, Justin Reich, Jonah Rockoff, Kirby Smith, and Oscar Sweeten-Lopez for providing helpful and detailed comments as we put together this review. We also thank Rachel Glennerster for detailed support throughout the project, Jessica Mardo and Sophie Shank for edits, and to the Spencer Foundation for financial support. Any errors or omissions are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

† Go to https://doi.org/10.1257/jel.20191507 to visit the article page and view author disclosure statement(s).
—have precipitated a renaissance in education technology (ed-tech), a term we use here to refer to any ICT application that aims to improve education. In the United States, the market for PreK–12 software alone has exceeded $8 billion (SIIA Communications 2015), and a recent industry report projects an estimated value of $252 billion for the global ed-tech industry by 2020 (Morrison 2017). Governments, schools, and families increasingly value technology as a central part of the education process, and invest accordingly (Bulman and Fairlie 2016). In the coming years, emerging fields like machine learning, big data, and artificial intelligence will likely compound the influence of these technologies even further, expanding the already dizzying range of available education products and speeding up cycles of learning and adjustment.

Collectively, these technologies offer the potential to open doors and build bridges by expanding access to quality education, facilitating communication between educators, students, and families, and alleviating frictions across a wide variety of educational contexts from early childhood through adulthood. For example, educational software developers work to enable educators to deliver the latest learning science advances to schools in inner cities and remote rural areas alike. The proliferation of cell phones and growing ease in connecting them to internet-based information systems has enabled the scaling of automated text messaging systems that aim to inform, simplify, and encourage students and their parents as they traverse difficult sticking points in education, like the transition from high school to college. And online educational institutions may bring opportunities to earn degrees to students who would otherwise be constrained by work, families, disabilities, or other barriers to traditional higher education. Taken together, technology may play a significant role in shaping the education production function, as technology has done in other sectors.

But the rapid proliferation of new technologies within education has proved to be a double-edged sword. The speed at which new technologies and intervention models are reaching the market has far outpaced the ability of policy researchers to keep up with evaluating them. The situation is well summarized by a recent headline: “Ed-Tech Surges Internationally—and Choices for Schools Become More Confusing” (Mohar 2017). While most agree that ed-tech can be helpful under some circumstances, researchers and educators are far from a consensus on what types of ed-tech are most worth investing in, in which contexts, and for which populations. At the same time, there is reason to believe that some uses of technology could potentially be harmful, and many parents are growing increasingly concerned that screen time may distract from student learning and development or that technology could displace the role of the teacher (Bowles 2019).

Furthermore, the transformations associated with ed-tech are occurring in a context of deep and persistent inequality. Despite expanding access to some technologies, the digital divide remains very real and very wide. While 98 percent of children in US households with incomes exceeding $100,000 per year have a computer at home, only 67 percent of children in households with incomes lower than $25,000 have them (Bulman and Fairlie 2016). Even when disadvantaged students can physically access technology, they may lack the guidance needed for productive utilization—a “digital-use divide” (Brotman 2016). Depending on design and implementation, education technologies could alleviate or aggravate existing inequalities. For at-risk students in particular, there are real concerns around the opportunity costs of receiving instruction through technology rather than through an in-person teacher or tutor. Equity considerations thus add another layer
to the need for caution when implementing technology-based education programs.

Of course, not every intervention model can be evaluated, and the extent of success inevitably varies across educational approaches and contexts even within well-established fields. But the speed and scale with which many ed-tech interventions are being adopted, along with the enormous impact they could have over the next generation, demand a closer look at what we know. To confront this issue, the present review takes stock of rigorous quantitative studies on technology-based education interventions that have been conducted so far, with the goals of identifying insights for the economics of education and highlighting key areas for future inquiry. In particular, for reasons explained in the following section, we assembled what we believe to be a comprehensive list of all publicly available studies on technology-based education interventions that report findings from studies following either of two research designs, randomized controlled trials or regression discontinuity designs, and based our analyses primarily on these studies.\footnote{For simplicity, and as explained further in the following section, we refer to studies employing these designs as “experimental research.”} The studies are clustered into four ed-tech domains, which generally focus on distinct families of tools, logics, and approaches: (i) access to technology, that is, opportunities to use computers, software, and other forms of ICT with the potential for use with education; (ii) computer-assisted learning (CAL), that is, computer programs and other software applications designed to improve academic skills; (iii) online courses, that is, classes offered by educational institutions partially or entirely through the internet; and (iv) behavioral interventions, that is, nudges, reminders, and other ICT-based programs that aim to overcome psychological or sociological barriers.

To position these findings within the broader economics literatures on education and human capital development, we organize them according to a framework based on Cunha and Heckman’s “technology of skill formation” model (Cunha et al. 2006, Cunha and Heckman 2007, see also figure 28 in Almlund et al. 2011). This model posits that the “skill formation process is governed by a multistage technology” spanning the educational life cycle, and that, at each stage, individuals “possess a vector of abilities…[that] are multiple in nature and range from pure cognitive…to noncognitive” (Cunha and Heckman 2007). An individual’s abilities and skills at a given stage are, in turn, a function of cognitive and noncognitive skills from the preceding stage in combination with environmental factors, and past investment in education or other forms of human capital. We define cognitive skills as the core skills that are used to think, read, and learn. On the other hand, noncognitive skills are traits such as personality, social, and emotional traits that are now increasingly acknowledged as being important contributors to learning.

Figure 1 illustrates how we adapt this framework to the ed-tech context. We narrow the focus from human capital in general to educational advancement in particular and focus on the role of ed-tech in improving environmental conditions or the efficiency of direct investments in cognitive or noncognitive skills at different life-cycle stages (early childhood, elementary, secondary, and postsecondary). Ed-tech consists of those technologies that rely on ICT. In this framework, ed-tech products are successful to the extent that they increase the efficiency of investment in skills relative to the next best technology (whether ICT-based or traditional) since “expenditures devoted to technology necessarily offset inputs that may be more or less efficient, and time allocated to using technology may displace traditional
classroom instruction and educational activities at home” (Bulman and Fairlie 2016). We hone in on the role technology may play in overcoming longstanding constraints in education.

Ed-tech products may address cognitive, noncognitive, or environmental factors at one or more educational life-cycle stages. The domains we examine tend to differ, however, with regard to the inputs they focus on in each stage. In particular, access-to-technology interventions tend to focus on improving environmental factors, while CAL tends to focus on cognitive skills. Online courses, on the other hand, may focus on both improving environmental factors—by expanding the environment where instruction is delivered—and supporting cognitive skill formation—by varying how content is taught. Behavioral interventions typically attempt to improve or fill in gaps caused by noncognitive skill deficits. By exploring the results of the experimental research on ed-tech in the context of this framework, we hope to shed light on how technology can resolve important frictions and binding constraints that impact the education production function and, ultimately, the accumulation of human capital.

In the next section, we discuss our literature review methodology in greater depth. Sections 3–6 constitute the core of the review, respectively synthesizing the evidence on access to technology, CAL, online courses, and behavioral interventions. In each of these sections, we discuss background and context for the ed-tech domain in question; consider how ed-tech programs within the domain seek to mitigate constraints and improve the efficiency of investments in cognitive skills, noncognitive skills, or environmental factors.

Figure 1. Technology of Skill Formation in the Education Technology Context

Source: Authors, inspired by Almlund et al. (2011), figure 28.
at each educational stage; review the experimental evidence on programs fitting into the domain; and reflect on implications for future research, theory, and policy. Section 7 offers concluding observations and considers several of the priority areas for future research that we consider vital to ongoing efforts at more effectively and equitably leveraging technology for learning.

2. Literature Review Methodology

While a few existing reviews have covered subsets of ed-tech, no recent review has attempted to cover the full range of ed-tech interventions (Bulman and Fairlie 2016; Lavecchia, Liu, and Oreopoulos 2016; Means et al. 2009). In particular, no previous review, to our knowledge, brings together computer- and internet-based learning on one hand with technology-based behavioral interventions on the other. We also focus on studies presenting evidence from randomized controlled trials (RCTs) and regression discontinuity designs (RDDs). Our core focus on RCT- and RDD-based studies constitutes a second unique contribution of this review—we argue that, in addition to helping us define sufficiently clear and narrow inclusion conditions, a focus on RCTs and RDDs adds a productive voice to broader and more methodologically diverse policy research dialogues in an environment characterized by complex tangles of cause and effect.

Why focus on RCTs and RDDs? In the field of applied microeconomics, RCTs—when properly implemented—are generally considered the strongest research design for quantitatively estimating average causal effects (Angrist and Pischke 2008). Because it is not always possible to conduct RCTs, methodologists have, over the past several decades, developed a toolkit of research designs such as instrumental variables, difference in differences, and propensity score designs, known broadly as quasi-experiments. RDDs are quasi-experiments that identify a well-defined cutoff threshold which defines a change in eligibility or program status for those above it—for instance, the minimum test score required for a student to be eligible for financial aid. While very high-scoring and very low-scoring students likely differ from one another in ways other than their eligibility for financial aid, “it may be plausible to think that treatment status is ‘as good as randomly assigned’ among the subsample of observations that fall just above and just below the threshold” (Lee and Card 2008). So, when some basic assumptions are met, the jump in an outcome between those just above and those just below the threshold can be interpreted as the causal effect of the intervention in question for those near the threshold (Imbens and Lemieux 2008, Thistlewaite and Campbell 1960). One caveat is that, while RCTs estimate the average treatment effect for the entire sample, RDDs measure the “local average treatment effect” (Imbens and Angrist 1994), that is, the treatment effect for cases near the threshold.

We chose to include RDDs, but not other quasi-experimental designs, because of their minimal sensitivity to underlying theoretical assumptions and the fact that RDDs with large samples and well-defined thresholds produce estimated program effects identical to RCTs for participants at the cutoff (Berk et al. 2010, Cook and Wong 2008, Shadish et al. 2011). Although RDDs are quasi-experiments, in the remainder of this review we refer to the RCTs and RDDs included in this review as experimental research for simplicity. We chose to focus on experimental research not only because we believe it can often be more convincing than other types of research for causal inference, but because we felt that the economic literature on ed-tech is flooded with observational research and could benefit from a synthesis of evidence from the designs most likely...
to produce unbiased estimates of causal effects. Furthermore, we introduce, frame, and interpret the experimental results in the context of broader observational literatures.

Experimental evaluation studies estimate the impact of a program or policy on outcomes of interest. But the estimates they come up with are sometimes difficult to compare with one another, given that studies test for impact on different outcomes using different measurement tools, in populations that differ in their internal diversity. While these differences cannot be completely eliminated and effect sizes must always be considered in the contexts within which they were identified, standard deviations offer a roughly comparable unit that can give us a broad sense of the general magnitude of impact across program contexts. We thus report effect sizes in standard deviations whenever the relevant data are available below in order to facilitate comparison, while cautioning that these effect sizes must be considered in context to be meaningful.

We also limited our core focus to studies conducted within developed countries, although we touch on research conducted in developing countries where relevant to the discussion. After considering both literatures, we determined that the circumstances surrounding the ed-tech interventions that have so far been experimentally studied differed too greatly across developed and developing country education systems to allow for integrating findings from both in a way that would yield meaningful policy implications. Our decision to focus on the developed rather than developing world in particular was driven by this review’s goal of analyzing experimental research on the full range of ed-tech interventions. While experimental policy and evaluation literature on certain classes of ed-tech literature, like computer distribution and CAL, have already begun to flourish in the developing world, experimental research on other areas, like technology-based behavioral interventions, is less developed there so far.

Our first task in constructing this review was thus to collect all publicly available studies using RCT or RDD designs within developed countries that estimate the effects of an ed-tech intervention on any education-related outcome. To locate the studies, we assembled a list of search terms, and used these to search a range of academic search engines, leading economics and education journals, and evaluation databases. To ensure that no relevant studies had been omitted, we followed backward and forward citations for all included articles.

The “file drawer” problem—the notion that studies with significant results are relatively more likely to be published, while studies showing null results tend to be filed away—presents a perennial challenge for literature reviews oriented toward impact evaluations. While no review can fully circumvent this challenge, we took steps to minimize its presence within this article. In particular, we chose not to exclude any studies based on publication status. Our final list thus consists of published academic articles, working papers, evaluation reports, and unpublished manuscripts. Furthermore, we conducted extensive consultations with leading researchers, evaluators, and practitioners in the field, asking each about every study that s/he was aware of in his or her area of specialization, whether or not the study was published or unpublished, and whether its findings were significant or null. The file drawer problem may extend beyond publication bias, in that papers may not even be written up if null results are detected. While we cannot entirely rule out this possibility, we believe we have taken all feasible steps to avoid it, and that our approach is certainly more effective in doing so than the modal literature review approach of solely performing keyword searches within databases that consist entirely of published articles. A related
problem is that researchers may emphasize the most significant findings in their studies such that interventions seem more effective than they actually are. We addressed this problem by closely reading each article and considering the broader constellations of estimates, rather than solely the headline findings, and integrating these considerations into our review as relevant.

Once the articles had been assembled, we divided them into the four categories into which we felt that they most naturally clustered: access to technology, CAL, technology-based behavioral interventions in education, and online courses. Although not all studies fit neatly into these categories and there is some overlap, we felt that these four best encapsulated the differences in the studies’ underlying themes, motivations, and theories of change. The full list of studies is contained—separated by category—in tables 1 through 4D.

Within each category, we closely read all studies and organized them further according to the approach of the intervention evaluated. We then considered each study’s findings in light of the others’, taking into account to the greatest extent possible variations in both the nature of the programs evaluated, the contexts in which they are implemented, and the research design specifics. Although we provide counts of the numbers of studies falling into different categories, this is done to provide readers with a view of the research landscape—we do not make inferences based on “vote counting,” that is, deciding on whether a particular intervention model is effective when there are more studies with positive than null or negative findings, or vice versa. Instead, we consider the big picture of research design, the nature of the intervention itself, and sample characteristics when putting forward substantive arguments. Where findings from the literature are ambiguous or inconclusive, we lay out central parameters along with the best of our interpretation and leave it to readers to make a final determination or to await future research. While quantitative techniques of meta-analysis have grown in popularity and sophistication in recent years, we felt that these techniques would not be appropriate for this paper because of our goals and the nature of the literature. Quantitative meta-analytic techniques work well when studies fall into relatively dense clusters with comparable outcomes and samples. However, with a few exceptions, the subliteratures reviewed here are relatively thin and internally diverse. And, as explained above, our goal was to bring together, within a single framework, the full range of ed-tech interventions that have been experimentally evaluated. Thus, we instead put forward a qualitative discussion, with the goals of integrating the studies into a broader economic framework, charting important findings and gaps, and distilling overarching lessons. In the remainder of the review, we present the results of this analysis.

3. Access to Technology

3.1 Background, Context, and Mechanisms of Impact

A natural starting point when exploring the effects of ed-tech is to consider what happens when students are provided with increased access to computers or the internet. Since the acceleration in ed-tech first began in earnest during the 1990s, governments and other stakeholders have invested substantial resources in an array of computer and internet distribution and subsidy initiatives. While the handful of studies that have been conducted so far on such initiatives are far from sufficient to judge the effectiveness of these efforts, they do suggest some early lessons. Access programs evaluated thus far have been effective at increasing use of computers and improving computer
skills. These outcomes alone are noteworthy, given the logistical challenges of technology distribution—particularly within lower-capacity and otherwise disadvantaged delivery contexts—and given the potential reluctance of students and educators to change their routines by incorporating the technologies. Effects on academic achievement and other learning outcomes, however, have been mixed at best. But the research suggests some potential areas of promise here as well, particularly computer distribution at the postsecondary level and distribution at the K–12 level when combined with additional learning software.

A large and growing share of students in developed countries can now access computers with high-speed internet at home and at school. Today, nearly three-quarters of American adults have broadband access at home—a remarkable increase from only one percent of adults in 2000 (Pew Research Center 2017). Among adults with children, the rate of at-home broadband access is even higher. A 2015 Pew Research Center study found that 82.5 percent of American households with school-age children have broadband access (Horrigan 2015).

But damaging holes in coverage remain. Approximately 5 million school-age children do not have a broadband internet connection at home (Anderson and Kumar 2017), potentially leading to a “homework gap” (Kang 2016) and other compounding layers of disadvantage. Students without computers or internet are likely to be the students who could most benefit from a boost in human capital, as they are much more likely to come from lower-income households: “In the United States, 98 percent of the 12 million schoolchildren living in households with $100,000 or more in income have access to a computer at home, but only 67 percent of the 12 million schoolchildren living in households with less than $25,000 in income have access” (Bulman and Fairlie 2016). And underrepresented minority students disproportionately lack access: only 78 percent of African American and Hispanic schoolchildren have computers at home, in contrast to 92 percent of white schoolchildren (Bulman and Fairlie 2016, p. 263). There is also a stark technology access divide between rural and urban areas (West and Karsten 2016).

Several program models have emerged to address these gaps in access to technology. One model that has recently risen to prominence has been “one-to-one” technology, “in which all the students in a class, grade level, school, or district are provided computers for use throughout the school day and, in some cases, at home” (Zheng et al. 2016). Several one-to-one initiatives have been implemented at large scales. For instance, the state of Maine provides all of its middle and high school students with laptops for use during the school year. More recently, some school districts around the country have been pairing students up with tablets (McLester 2012). One-to-one distribution has also caught on within developing countries, and governments as diverse as those of Peru, Kenya, Turkey, and India have invested in variations of such programs (Trucano 2013, BBC 2013, Simhan 2011). One particularly prominent civic-led one-to-one initiative has been the One Laptop per Child (OLPC) program, which aims to “empower the children of developing countries to learn by providing one connected laptop to every school-age child.” OLPC has distributed laptops to disadvantaged students in roughly a dozen developing countries, along with two US cities.

Other initiatives have provided schools with subsidies to buy computers or software,
or to acquire or improve internet connections. In 1997, the US federal government launched its largest-ever program to connect US schools and classrooms to the internet. Known as E-Rate, the program has connected 97 percent of US classrooms to the internet. In 2013, President Barack Obama announced a new initiative known as ConnectED, which sought to bring high-speed broadband to 99 percent of K–12 students by 2018 (Benton Institute for Broadband & Society 2013). The initiative helped provide an additional 20 million students with in-classroom access to broadband. Both the private and public sectors have invested heavily to increase broadband access around the country. Since 2009, more than 115,000 miles of network infrastructure have been built at a cost of more than $260 billion (Council of Economic Advisers 2016).

Unlike the other ed-tech domains discussed in this article, interventions focusing on access to technology do not directly seek to alter the skill formation process itself. Instead, as illustrated by their placement within our framework, these interventions aim to help create an environment that is more favorable to skill formation. As an increasing share of education in and out of the classroom occurs through ICT, availability of the hardware and software through which ICT operates becomes increasingly vital to children's and students’ educational environments. In particular, availability of ICT may overcome traditional constraints associated with how we access information and how quickly we can access information. For example, rather than going to the library to check out a book, many students today can simply download a book from the internet, reducing time costs associated with doing homework and freeing up time and resources for other learning activities. Much as access to books represented an important feature of the educational environment throughout the twentieth century, access to computers and internet will likely represent an important feature of educational environments in the twenty-first century. On the other hand, access to computers and the internet could harm learning outcomes if they lead students to spend time on noneducational games or other leisure activities that take time away from studying.

### 3.2 Experimental Evidence

While access to technology is clearly a necessary condition for the operation of ed-tech, the extent to which technology access programs result in measurably improved outcomes remains an empirical question. Given the wave of investment and policy interest in access to technology, what have been the effects of access programs? With only a handful of RCT and RDD studies on the subject, the experimental literature on its own cannot say much definitively. However, these studies provide valuable suggestive insights, particularly when viewed within the context of the broader quasi-experimental and observational literatures.

In particular, we identified 13 experimental papers that we coded as studying access to technology interventions—nine reporting on RCTs, and four reporting on RDDs. These papers are laid out in Table 1. Of the nine papers reporting on RCTs, eight are authored or coauthored by economist Robert W. Fairlie and are based on two impact evaluations. Three of the papers (Fairlie and Robinson 2013, Fairlie and Kalil 2017) report findings from a relatively large-scale impact evaluation of a laptop distribution program for sixth to tenth graders, with a sample of 1,123 students in 15 schools covering five school districts across California. The remaining five (Fairlie 2012a, b; Fairlie and Bahr 2018; Fairlie and London 2012;
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Standardized effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carter, Greenberg, and Walker (2017)</td>
<td><em>Economics of Education Review</em></td>
<td>Prohibiting use of computers during a college economics class</td>
<td>Positive effect on final exam scores</td>
<td>0.18 standard deviations</td>
<td>50 postsecondary classrooms and 726 students in West Point, New York</td>
<td>726 students in 50 classrooms</td>
<td>Classroom</td>
</tr>
<tr>
<td>Faber, Sanchis-Guarner, and Weinhardt (2015)</td>
<td>NBER Working Paper</td>
<td>Differences in broadband connection speeds</td>
<td>Null effects on student time spent studying or learning productivity</td>
<td>Null</td>
<td>Primary and secondary school students living in residential postcodes in England</td>
<td>4.5 million students in 580,000 postcodes in England</td>
<td>Not randomized—spatial regression discontinuity using postcodes within 1 km of telephone exchange system boundaries</td>
</tr>
<tr>
<td>Fairlie (2012a)</td>
<td><em>Economics of Education Review</em></td>
<td>One-to-one laptop distribution</td>
<td>1. Null effect on GPA 2. Positive impact on course completion rates and course success rate 3. Null effect on graduation rate</td>
<td>Main reported effect sizes (not standardized): 1. Null on GPA 2. 6.5 percentage point difference in course completion rates 3. 8.6 percentage point difference for course success rate</td>
<td>Community college students receiving financial aid in California</td>
<td>286 students</td>
<td>Student</td>
</tr>
<tr>
<td>Fairlie (2012b)</td>
<td><em>Information Economics and Policy</em></td>
<td>One-to-one laptop distribution</td>
<td>Positive effect on high-level computer skills, especially among young, minority, low-income, and female students</td>
<td>Main reported effect sizes (not standardized): 17 percentage point difference</td>
<td>Community college students receiving financial aid in California</td>
<td>286 students</td>
<td>Student</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Standardized effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairlie (2015)</td>
<td>B.E. Journal of Economic Analysis &amp; Policy</td>
<td>One-to-one laptop distribution</td>
<td>Null results on academic outcomes</td>
<td>Null</td>
<td>Children enrolled in grades 6–10 in 15 different middle and high schools in 5 school districts in California</td>
<td>1,123 students</td>
<td>Student</td>
</tr>
<tr>
<td>Fairlie and Bahr (2018)</td>
<td>Economics of Education Review</td>
<td>One-to-one laptop distribution</td>
<td>Null results on earnings</td>
<td>Null</td>
<td>Community college students receiving financial aid in California</td>
<td>286 students</td>
<td>Student</td>
</tr>
<tr>
<td>Fairlie and Grunberg (2014)</td>
<td>Economic Inquiry</td>
<td>One-to-one laptop distribution</td>
<td>Positive effect on transferable courses enrollment; null results on actual transfer to 4-year colleges</td>
<td>Main reported effect sizes (not standardized): 1. 4.5 percentage point difference on enrollment in courses that are transferable to 4-year colleges 2. Null</td>
<td>Community college students receiving financial aid in California</td>
<td>286 students</td>
<td>Student</td>
</tr>
<tr>
<td>Fairlie and Kalil (2017)</td>
<td>Economics of Education Review</td>
<td>One-to-one laptop distribution</td>
<td>1. Null effect on social development 2. Positive effect on online social networking 3. Positive effect on in-person friend interaction</td>
<td>1. Null 2. 9 percentage point increase in having an online social networking site (not standardized) 3. 0.10 standard deviation increase in friends outcome index</td>
<td>Children enrolled in grades 6–10 in 15 different middle and high schools in 5 school districts in California</td>
<td>1,123 students</td>
<td>Student</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Standardized effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>--------------------------</td>
<td>--------</td>
<td>-------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Fairlie and London (2012)</td>
<td><em>Economic Journal</em></td>
<td>One-to-one laptop distribution</td>
<td>Positive effect on a “summary index of educational outcomes” that includes variables like grades and degree completion, especially for students who live farther from campus or have a job</td>
<td>0.14 standard deviations</td>
<td>Community college students receiving financial aid in California</td>
<td>286 students</td>
<td>Student</td>
</tr>
<tr>
<td>Fairlie and Robinson (2013)</td>
<td><em>American Economic Journal: Applied Economics</em></td>
<td>One-to-one laptop distribution</td>
<td>Null results</td>
<td>Null</td>
<td>Sixth–tenth graders in 15 middle and high schools in 5 districts in California; vast majority of sample is middle school students</td>
<td>1,123 students</td>
<td>Student</td>
</tr>
<tr>
<td>Goolsbee and Guryan (2006)</td>
<td><em>Review of Economics and Statistics</em></td>
<td>E-Rate, subsidy for internet in schools</td>
<td>Null results on academic outcomes; Positive effect on internet connectivity</td>
<td>Main reported effect sizes (not standardized): 1. Null on academic outcomes 2. 68 percent increase in internet-connected classrooms per teacher</td>
<td>California public schools (primary, middle, and high schools)</td>
<td>31,240 classrooms with internet access</td>
<td>Not randomized—regression discontinuity-based design</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Standardized effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leuven et al. (2007)</td>
<td><em>Review of Economics and Statistics</em></td>
<td>Subsidies for computers and software in under-resourced schools</td>
<td>Null results</td>
<td>Main reported effect sizes (not standardized): 1. Null (seems especially detrimental for girls)</td>
<td>Primary schools in the Netherlands that had at least 70 percent of pupils belonging to the disadvantaged minority</td>
<td>267 schools in 1998, and 551 schools in 1999</td>
<td>Not randomized—regression discontinuity design</td>
</tr>
<tr>
<td>Malamud and Pop-Eleches</td>
<td><em>Quarterly Journal of Economics</em></td>
<td>Euro 200 program, subsidy for low-income families with schoolchildren to buy computers</td>
<td>1. Negative effect on math/English/Romanian 2. Positive effect on computer skills 3. Positive effect on cognitive skills (not robust in all specifications)</td>
<td>1. 0.25–0.33 standard deviation decrease in math/English/Romanian 2. 0.27 standard deviation increase in computer skills 3. 0.33 standard deviation increase in cognitive skills</td>
<td>Households from several regions of Romania (primary and secondary students)</td>
<td>3,354 households</td>
<td>Not randomized—regression discontinuity design</td>
</tr>
</tbody>
</table>
Fairlie and Grunberg 2014) report on a smaller evaluation of a laptop distribution program in a single California community college with a sample of 286 students on financial aid. Another field experiment estimates the effect of prohibiting laptop use during class time in a West Point economics course (Carter, Greenberg, and Walker 2017), and the remaining four are RDD studies estimating the impact of access to computers (Leuven et al. 2007, Malamud and Pop-Eleches 2011) or internet (Faber, Sanchis-Guarner, and Weinhardt 2015; Goolsbee and Guryan 2006) on a variety of education-related outcomes.

Despite the differences in interventions and settings explored within the studies, the papers consistently report success in programs’ intended proximate outcomes—distributing computers, increasing time spent using computers, or decreasing time spent accessing computers (e.g., less time waiting for computers in labs to become available). For example, in the California field experiment—the only large-scale randomized evaluation to date of laptop distribution programs—computer ownership reportedly increased by 55 percentage points among treatment students, computer usage reportedly increased by 2.5 hours per week, and the likelihood of at-home internet connection increased by 25 percentage points relative to those who were not assigned to receive free laptops (Fairlie and Robinson 2013). These findings are noteworthy considering that the significant resources required to expand computer and internet access may be wasted because of the logistical difficulties of distribution. And students and teachers facing constraints on time and cognitive capacity may be reluctant to adopt technologies in the ways intended by providers.

On the other hand, findings of effects on learning outcomes have been more mixed, with little evidence to suggest that technology access alone will improve education. The study on laptop distribution in California found precisely estimated null effects on learning outcomes. In particular, no significant impact—positive or negative—was found on homework time, grades, standardized test scores, attendance, or several other outcomes, leading the authors to conclude that “increasing access to home computers among students who do not already have access is unlikely to greatly improve educational outcomes, but is also unlikely to negatively affect outcomes” (Fairlie and Robinson 2013). At the same time, the estimated average effects may be masking heterogeneous outcomes, with some students benefiting from receiving access to computers and others doing worse. Perhaps even more concerning are the findings from an RDD study of subsidized computers for households in Romania (Malamud and Pop-Eleches 2011) and another that subsidized schools in purchasing computers and software in the Netherlands (Leuven et al. 2007). Both found negative impacts on achievement outcomes, likely in part a result of the students spending more time playing games. However, the negative effects in the Netherlands study are minor, and in the Romania study negative impacts on academic achievement are accompanied by positive impacts on computer skills and cognitive test scores. RDD studies that respectively looked at effects of the E-Rate school internet subsidy program in the United States (Goolsbee and Guryan 2006) and coincidental differences in connection speed in England (Faber, Sanchis-Guarner, and Weinhardt 2015) similarly found no evidence of substantial positive or negative impact on academic achievement.

On the other hand, one randomized evaluation of laptop distribution for students on financial aid in a community college finds positive results on academic outcomes. As reported in five papers, this study observed modest but positive effects, with
an overall impact on an academic performance index of 0.14 standard deviations (Fairlie and London 2012). The academic performance index is a measure the authors constructed to aggregate four separate outcomes: course success rate, the likelihood of taking a course for a grade, the likelihood of taking a transfer course for a four-year college, and graduation rate. Analyses suggested that the benefits occurred not by increasing the time that students spend using computers, but by reducing costs associated with time using computers in the college’s computer labs. Two separate papers reporting on the same study also find that positive academic effects are significantly stronger for minority than for nonminority students (Fairlie 2012a), and that the program increased computer skills most strongly for minorities, women, lower-income, and younger students (Fairlie 2012b). However, a follow-up study showed no impact on earnings seven years after the program was implemented (Fairlie and Bahr 2018).

With only two randomized evaluations (albeit reported on in nine papers) and four disparate RDD studies, the experimental literature in the developed world on its own cannot speak definitively to the effectiveness of programs seeking to improve access to technology and the internet. However, the consistency of findings of null impact on learning outcomes for elementary and secondary students—and especially the relatively large-scale and well-powered study of the laptop distribution program among sixth–tenth graders in California—indicates that efforts to design and test technology interventions may do best to focus elsewhere. On the other hand, the positive findings in the community college study by no means prove the potential of laptop distribution at the community college level, particularly given the evaluation’s small sample size. But given that effects seem to have worked through reduction in time spent waiting to use computers may suggest that, to the extent that there are resources for expanding access to technology, it may be most effective to invest them in contexts where the technology is clearly integrated into an educational curriculum, but where accessing the technology is somehow costly to students.

Where do these findings from the experimental literature in developed countries stand within the broader literature on interventions related to technology access? Relative to the other sections, technology access interventions have been well studied within the developing world. This research has, for the most part, come up with similar results as for the developed world. Interventions giving computers to schools in Colombia6 One Laptop per Child efforts in Peru (Beuermann et al. 2015, Cristia et al. 2017), and tablets distributed to students in Kenya (Piper et al. 2016) showed no impact on learning outcomes in the experimental studies. One study based in rural Peru reported positive effects on cognitive outcomes and another study based in China (Mo et al. 2015) found that increased access to computers improved math scores. Perhaps instructively, the intervention in China was the only one of the computer distribution initiatives in which computers were reliably equipped with educational software that was actually used by the students, which suggests that access to technology may be effective when there is clear integration with a broader educational program. When viewed within our framework, this study suggests that initiatives may be effective when they attempt to improve the environment and cognitive skill formation simultaneously.

Observational and quasi-experimental studies in both developed and developing

---

6Barrera-Osorio and Linden (2009), Rodriguez et al. (2011) find a positive impact from the same program after more time had elapsed, but the latter study is primarily nonexperimental.
countries have tended to find more positive results. One recent review of both observational and experimental studies on one-to-one programs implemented between 2001 and 2015 finds that an expansive range of positive impacts have been documented, including “…increased academic achievement in science, writing, math, and English; increased technology use for varied learning purposes; more student-centered, individualized, and project-based instruction; enhanced engagement and enthusiasm among students; and improved teacher–student and home–school relationships” (Zheng et al. 2016). However, many of the studies reviewed are not equipped for rigorous causal inference. These conclusions should thus be tempered in light of the experimental findings discussed in this review, which suggests that increased access to technology may create a better environment by reducing constraints associated with acquiring information or completing homework, but by itself may not necessarily facilitate more efficient formation of cognitive or noncognitive skills.

3.3 Looking Forward

What insights does the experimental literature bring to ongoing debates on ed-tech? First, more research is needed on efforts to improve access to technology at the postsecondary level. Postsecondary education demands a variety of more complex tasks that, in many cases, truly necessitate the need for a computer. Although students enrolled in colleges are more likely to already have computer access (Anderson 2015, School Guides 2014, Advanced Micro Devices 2014), computer ownership and internet access are far from universal among lower-income and otherwise disadvantaged students (Cohn 2014) and seeking access to computers at labs may waste scarce time or other resources. Notwithstanding the lack of impact found on earnings, the one randomized evaluation on this topic has shown promising results in this area, but findings from a single study at a single college may be difficult to generalize. More research is needed to understand the long-term impacts of increased access to technology especially given technology’s outsize role in tomorrow’s labor market. If we believe that increased computer proficiency is related to one’s ability to succeed in an increasingly automated workforce, increased access to technology during childhood may support human capital development in ways not captured by the existing research.

Second, while the few technology access programs that have been experimentally evaluated at the primary and secondary levels show few positive effects on academic achievement, improving access in combination with other activities may yield better results. For instance, the survey conducted for the Romania study discussed above found some suggestive evidence that the negative effects of home computers on grades was attenuated with certain parental rules—approaches to regulating children’s computer use or providing more structure and guidance for how the computer should be used may be worth studying. More research is needed to understand the interaction between increased access to technology coupled with learning-oriented activities. And, although increasing access to computers and internet may not on their own measurably improve academic achievement, they have been successful in increasing the time and/or ease of use. This observation, in combination with the positive results found for educational software discussed in the following section, suggests that the most promising policy models may be those that integrate hardware distribution with particular learning programs. We turn to discussing learning software programs in the following section.
4. Computer-Assisted Learning

4.1 Background, Context, and Mechanisms of Impact

Computer and learning scientists have been working for decades to develop software to deliver educational content, and the popularity of these programs has exploded in the wake of the 1990s’ ICT revolution. For the purposes of this review, we refer to initiatives relating to educational software as CAL programs. CAL programs differ from the technology distribution programs of the previous section in that they do not involve the provision of hardware for general use, but instead center on “well-defined” (Rouse and Krueger 2004) uses of specific software packages. Moreover, they differ from the online courses discussed in the following section in that they are software packages designed to develop particular skills, for example, improving math computation or improving reading comprehension, rather than platforms through which to administer courses. A multitude of companies have entered the market to meet growing demand from educators and policy makers for CAL, resulting in the advent of a plethora of products being used daily by millions of students worldwide.

Under what circumstances might we expect CAL to deliver significant impacts on student learning? In terms of our framework, CAL programs generally function as inputs that aim to improve the formation of cognitive skills. With this in mind, a CAL program is successful to the extent that it improves formation of cognitive skills more so than the next best technology—or, for practical purposes, business as usual. One way CAL products may supplement the skill formation process is by overcoming binding constraints faced by teachers related to managing heterogeneous learning levels within a single classroom.

Perhaps the most central mechanism through which CAL is expected to personalize learning and overcome traditional classroom constraints is adaptivity—the increasingly sophisticated ability of software to harness emerging artificial intelligence and machine-learning techniques to model the cognitive processes of students and to offer content accordingly.

An exciting body of experimental studies has reported large and significant effects from instructional models that enable students to dedicate time working through content that matches their level of academic preparedness (Banerjee et al. 2007, Banerjee et al. 2016). Such efforts can better allow students to master relatively basic concepts before moving on to more advanced concepts, and to practice more in areas where they are struggling and less in areas in which they are already strong.

Yet when teaching a classroom with students of many different levels, even the most skilled teacher can only adapt so much—a major constraint that educators have long attempted to overcome. When confronted with a wide range of student ability, teachers often end up teaching the core curriculum and tailoring instruction to the middle of the class. By leveraging technology, CAL programs have the potential to overcome these long-standing instructional constraints and amplify cognitive skill formation by efficiently targeting content to every student.

Aside from directly tailoring content toward students, CAL programs can overcome significant time constraints that teachers face and enhance the efficiency of an
educator’s efforts to facilitate cognitive skill formation in his or her students. For example, CAL programs can offer students individualized feedback and rapidly gather data on student performance, which would be challenging for any educator to do given time constraints. The program theories guiding the interventions evaluated in the studies that we review typically include multiple approaches to improving language, math, and other cognitive skills. While many CAL programs attempt to improve education by facilitating the increased personalization of learning, these programs vary widely in how they do so. They can range from light-touch interventions that provide practice opportunities outside of class, to more intensive interventions that provide courses with entirely new curricula, to (in a few cases) initiatives in which schools are organized entirely around CAL.

4.2 Experimental Evidence

To what extent and under what circumstances are these CAL programs effective? In this subsection, we review the experimental literature on this question. We identified 31 experimental studies of CAL programs in developed countries, all based on RCTs. While CAL can conceivably include a wide range of program types from games to research and networking tools, the CAL programs that have been evaluated experimentally generally fall within the broad category of “intelligent tutoring systems,” that is, software systems that aim to help students practice particular skills (Kulik and Fletcher 2015). As a caveat, the types of CAL programs that have been experimentally evaluated constitute only a small portion of the many CAL products on the market and currently used by schools, and the results of these studies cannot be generalized to other CAL products that might not function within similar theories of change. Nevertheless, taken together, the findings from these studies suggest that CAL programs of the types evaluated in these studies show enormous promise in improving learning outcomes, particularly when it comes to mathematics. Of the 31 studies included, 21 report statistically significant effects, with many precisely estimated and substantial in magnitude. The majority of the studies finding positive effects (16 of 21) were focused on improving math outcomes.\footnote{Barrow, Markman, and Rouse (2009); Beal et al. (2013); Hegedus, Dalton, and Tapper (2015); Karam et al. (2017); Kelly et al. (2013); Morgan and Ritter (2002); Pane et al. (2014); Ragosta, Holland, and Jamison (1982); Ritter et al. (2007); Roschelle et al. (2010); Roschelle et al. (2016); Schenke, Rutherford, and Farkas (2014); Singh et al. (2011); Snipes et al. (2015); Tatar et al. (2008); Wang and Woodworth (2011). Pane (2014) only finds positive impacts on math outcomes in the second year. Campuzano et al. (2009) did not focus exclusively on math outcomes and is therefore not included in this count.}

Math content may be particularly suited to personalized learning software that matches students’ learning levels since answers to math problems are typically more objective relative to language. Hence, software designers can base algorithms on formal rubrics that judge correct and incorrect answers to reliably place students at specific learning levels. Information on these studies is presented in Table 2.

Of those evaluated, several interventions show especially strong promise, for example, an evaluation of a math homework program in Maine showed an effect size of 0.18 standard deviations despite involving less than 30–40 minutes per week (Roschelle et al. 2016), while a more intensive software-based math curriculum intervention in Texas improved seventh and eighth grade math scores by 0.63 and 0.56 standard deviations, respectively (Roschelle et al. 2010). Many of the CAL interventions compare favorably with interventions commonly discussed in the economics of education literature like reduced class sizes, longer school days, and intensive face-to-face tutoring. We first review findings from studies on CAL.
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrow, Markman, and Rouse (2009)</td>
<td><em>American Economic Journal: Economic Policy</em></td>
<td>I Can Learn© also known as “Interactive Computer Aided Natural Learning” program for pre-algebra and algebra</td>
<td>Positive effect on test scores</td>
<td>0.17 standard deviations increase on a pre-algebra and algebra test; the strongest effects were for larger classes (especially with more heterogeneity in student levels) and classes with more absences</td>
<td>8 high schools and 2 middle schools in 3 large urban school districts in the Northeast, Midwest, and South with a high proportion of minority students</td>
<td>1,605 students in 142 classes</td>
<td>Classroom</td>
</tr>
<tr>
<td>Beal et al. (2013)</td>
<td>Presented at annual meeting of the American Educational Research Association</td>
<td>AnimalWatch web-based math tutoring program</td>
<td>Positive effects on student scores</td>
<td>0.3 standard deviations (approximate)</td>
<td>58 teachers’ classes for sixth grade</td>
<td>1,291 students in 58 teachers’ classes</td>
<td>Teacher</td>
</tr>
<tr>
<td>Borman, Benson, and Overman (2009)</td>
<td><em>Educational Evaluation and Policy Analysis</em></td>
<td>Fast ForWord computer-based language and reading training program</td>
<td>Null results for second graders in reading and language and for seventh graders in language; positive effects on reading comprehension for seventh graders</td>
<td>1. Second graders: null in reading comprehension or language 2. Seventh graders: 0.21 standard deviation increase in reading comprehension and null in language</td>
<td>Second and seventh grade students in Baltimore in 8 elementary and middle schools</td>
<td>415 students</td>
<td>Student</td>
</tr>
<tr>
<td>Cabalo, Ma, and Jaciw (2007)</td>
<td><em>Empirical Education</em></td>
<td>Cognitive Tutor’s Bridge to Algebra program</td>
<td>Null results</td>
<td>Null</td>
<td>32 pre-algebra classes in 5 schools in the Maui, Hawaii, school district</td>
<td>32 classrooms</td>
<td>Classroom</td>
</tr>
</tbody>
</table>

(Continued)
TABLE 2
COMPUTER-ASSISTED LEARNING (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
</table>
| Campuzano et al. (2009)   | Institute of Education Sciences | 16 types of software products for math and reading                           | 1. Null results for reading  
2. Negative results for sixth grade math scores  
3. Positive results for Algebra I test scores | 1. Null for reading  
2. Lower sixth math test scores in the second year than in the first year  
3. 0.15 standard deviation increase in Algebra I student test scores in the second year over the first year. | 33 US school districts, 132 schools, 428 teachers. It focused on school districts that had low student achievement and large proportions of students in poverty. | 9,424 students in 439 teachers’ classes in first cohort, 3,280 students in 176 teachers’ classes in second cohort | Teacher               |
| Cavalluzzo et al. (2012)* | Institute of Education Sciences | Kentucky Virtual Schools hybrid program for Algebra 1 | Null results | Null | 47 Kentucky schools (30 of which were in rural areas) with grade 9 algebra classes | 6,908 students in 25 schools | School | |
| Dynarksi et al. (2007)    | Institute of Education Sciences | 16 types of software products for math and reading                           | Null results | Null | 33 US districts, 132 schools, and 439 teachers participated in the study; it focused on school districts that had low student achievement and large proportions of students in poverty | 9,424 students in 439 teachers’ classes | Teacher | |

*This could also be a blended online learning and face-to-face intervention. In Kentucky Virtual Schools, instruction time is 60 percent face-to-face instruction and 40 percent is using online resources. The findings from this paper are consistent with the outcomes we observe in other blended classroom interventions.

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deault, Svage, and Abrami (2009)</td>
<td><em>Journal of Research on Educational Effectiveness</em></td>
<td>ABRACADABRA web-based literacy program</td>
<td>Positive and null effects in reducing negative correlations between attention and learning outcomes</td>
<td>1. Synthetic group: 0.41 standard deviations on listening comprehension, null on vocabulary, and 0.35 standard deviations on reading comprehension; null and positive impacts across more than a dozen other reading and related measures 2. Analytic group: null on listening comprehension, vocabulary, and reading comprehension; mainly null across more than a dozen other reading and related measures</td>
<td>Grade 1 students from schools in Montreal, Canada, from 13 different classrooms.</td>
<td>144 students</td>
<td>Students</td>
</tr>
<tr>
<td>Faber and Visscher (2018)</td>
<td><em>Computers &amp; Education</em></td>
<td>Snappet digital formative assessment tool focused on spelling</td>
<td>Null impacts on spelling achievement</td>
<td>Null</td>
<td>69 Dutch primary schools</td>
<td>1,605 students in 69 schools</td>
<td>School</td>
</tr>
<tr>
<td>Hegedus, Dalton, and Tapper (2015)</td>
<td><em>Educational Technology Research &amp; Development</em></td>
<td>SimCalc interactive math software</td>
<td>Positive effect on student learning of core algebra concepts</td>
<td>1. 0.29 standard deviations 2. 0.25 standard deviations in replication study</td>
<td>7 high schools in Southeast Massachusetts of varying achievement levels</td>
<td>280 students in 30 classrooms</td>
<td>Classroom</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------</td>
<td>--------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Karam et al. (2017)</td>
<td>Educational Technology Research &amp; Development</td>
<td>Cognitive Tutor Algebra I</td>
<td>Positive effect on Algebra 1 outcomes</td>
<td>0.14 increase from nontraditional student activities (use of traditional student activities was negatively associated with student outcomes on Algebra I for middle schools in years 1 and 2).</td>
<td>Middle and high schools in 51 school districts representing 7 states in the US that varied in contexts.</td>
<td>469 teachers in 73 high schools and 74 middle schools</td>
<td>School</td>
</tr>
<tr>
<td>Kelly et al. (2013)</td>
<td>Artificial Intelligence in Education</td>
<td>ASSISTments online math homework support</td>
<td>Positive effects on learning for students and teacher’s review of homework</td>
<td>0.56 standard deviations</td>
<td>Seventh grade students who were currently enrolled in an eighth grade math class in a suburban middle school in Massachusetts</td>
<td>63 students</td>
<td>Student</td>
</tr>
<tr>
<td>Mitchell and Fox (2001)</td>
<td>Reading Research and Instruction</td>
<td>DaisyQuest and Daisy’s Castle reading game</td>
<td>1. Positive effects on phonological processing  2. The teacher-facilitated computer instruction group outperformed the computer-administered group on several literacy measures.</td>
<td>1. 0.69 standard deviation increase in phonological awareness from computer-administered phonological instruction  2. 1.03 standard deviation increase in phonological awareness from teacher-facilitated computer instruction</td>
<td>US kindergarten and first grade students</td>
<td>72 students</td>
<td>Student</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td>-------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Morgan and Ritter</td>
<td>Unpublished</td>
<td>Cognitive Tutor</td>
<td>Positive effects in math outcomes on ETS test</td>
<td>0.29 standard deviations</td>
<td>Ninth graders in 5 junior high schools in Moore Independent School District, Oklahoma</td>
<td>444 students</td>
<td>Students</td>
</tr>
<tr>
<td>Pane et al. (2010)</td>
<td><em>Journal of Research on Educational Effectiveness</em></td>
<td>Cognitive Tutor</td>
<td>1. Negative effects on math outcomes 2. No effect on student attitudes toward mathematics and technology.</td>
<td>1. –0.19 standard deviations 2. Null</td>
<td>8 high schools in Baltimore Country Public School District (BCPS)</td>
<td>699 students in 19 teachers' classes</td>
<td>Teacher and student</td>
</tr>
<tr>
<td>Pane et al. (2014)</td>
<td><em>Educational Evaluation and Policy Analysis</em></td>
<td>Cognitive Tutor</td>
<td>1. No effect in the first year 2. Positive effect in second year for high schoolers</td>
<td>1. Null 2. 0.20 standard deviations in second year for high schoolers, but null for middle schools.</td>
<td>Public middle and high schools across 7 US states and 51 school districts in urban, suburban, and rural areas</td>
<td>18,700 high school students in 73 high schools, 6,800 middle school students in 74 middle schools</td>
<td>School</td>
</tr>
<tr>
<td>Ragosta, Holland, and Jamison (1982)</td>
<td>Unpublished (report available through Educational Testing Service)</td>
<td>Cognitive Tutor for math</td>
<td>Positive effects on student test scores in math and reading curriculum specific tests</td>
<td>1. Math: 0.80 standard deviations in year 1; 0.91 standard deviations in year 2; and 1.23 standard deviations in year 3 2. Reading: 0.38 standard deviations in year 1; 0.52 standard deviations in year 2; and null in year 3</td>
<td>Four elementary schools in Los Angeles</td>
<td>Unknown</td>
<td>Students</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ritter et al. (2007)</td>
<td>Supporting Learning, Flow through Integrative Technologies</td>
<td>Cognitive Tutor for math</td>
<td>Positive effect on student math grades; null on Algebra ETS score</td>
<td>0.416 standard deviations for first semester grades; 0.356 standard deviations for final grades; null for Algebra ETS score</td>
<td>Ninth graders in 5 junior high schools in Moore Independent School District, Oklahoma</td>
<td>343 students in 19 class sections</td>
<td>Class sections</td>
</tr>
<tr>
<td>Rockoff (2015)</td>
<td>Unpublished</td>
<td>School of One middle school mathematics program</td>
<td>Null results</td>
<td>Null</td>
<td>New York City public schools</td>
<td>5,070 students in 8 schools</td>
<td>School</td>
</tr>
<tr>
<td>Roschelle et al. (2010)</td>
<td>American Educational Research Journal</td>
<td>SimCalc interactive math software</td>
<td>Positive effects on math outcomes</td>
<td>0.63 standard deviations for seventh grade classrooms and 0.50 standard deviations for eighth grade classrooms</td>
<td>Seventh and eighth grade classrooms in Texas public schools</td>
<td>140 seventh grade classrooms in 77 schools, 80 eighth grade classrooms in 56 schools</td>
<td>School</td>
</tr>
<tr>
<td>Roschelle et al. (2016)</td>
<td>AERA Open</td>
<td>ASSISTmants online math homework support</td>
<td>Positive effects on math outcomes</td>
<td>0.18 standard deviations</td>
<td>Seventh graders in Maine</td>
<td>2,850 students in 43 schools</td>
<td>Schools</td>
</tr>
<tr>
<td>Rouse and Kneuger (2004)</td>
<td>Economics of Education Review</td>
<td>Fast ForWord computer-based language and reading training program</td>
<td>Null results</td>
<td>Null</td>
<td>4 schools in an urban school district in the Northeast; around 40 percent African American and 50 percent Hispanic</td>
<td>89 students</td>
<td>Students</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schenke, Rutherford, and Farkas (2014)</td>
<td>Computers &amp; Education (ST) Math</td>
<td>Spatial–Temporal Math</td>
<td>Positive effects on basic number sense skills</td>
<td>0.14 standard deviations on the Number Sense I (NSI) skills test</td>
<td>50 elementary schools in Southern California</td>
<td>10,860 students from 50 schools</td>
<td>School</td>
</tr>
<tr>
<td>Singh et al. (2011)</td>
<td>Unpublished</td>
<td>ASSISTments online math homework support</td>
<td>Positive effects on math outcomes</td>
<td>0.40 standard deviations</td>
<td>8 classes of eighth grade students in Maine</td>
<td>172 students</td>
<td>Student</td>
</tr>
<tr>
<td>Snipes et al (2015)</td>
<td>Institution of Education Sciences</td>
<td>Elevate summer math program</td>
<td>Positive effects on math achievement and algebra readiness, yet still did not prepare students for Algebra I content</td>
<td>0.70 standard deviations on a test of algebra readiness</td>
<td>Eighth grade students from 8 schools in 6 districts in California's Silicon Valley</td>
<td>496 students</td>
<td>Student</td>
</tr>
<tr>
<td>Tatar et al. (2008)</td>
<td>Journal of the Learning Sciences</td>
<td>SimCalc interactive math software</td>
<td>Positive effects on student and teacher mathematics knowledge</td>
<td>Main reported effect size (not standardized): positive effects on student and teacher mathematics knowledge (exact magnitude not reported)</td>
<td>21 seventh grade mathematics teachers in Texas</td>
<td>827 students in 21 teachers’ classes</td>
<td>Teacher</td>
</tr>
<tr>
<td>Van Klaveren, Vonk, and Cornelisz (2017)</td>
<td>Economics of Education Review</td>
<td>Adaptive CAL program compared against a static one across multiple subjects</td>
<td>Null results</td>
<td>Null</td>
<td>Dutch secondary schools</td>
<td>1,021 students</td>
<td>Student</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang and Woodworth (2011)</td>
<td>SRI International Center for Education Policy</td>
<td>1. DreamBox math program 2. Reasoning Mind math program</td>
<td>1. Positive effects from DreamBox on test scores 2. Null results from Reasoning Mind on test scores</td>
<td>1. DreamBox: 0.14 standard deviations on NWEA math test and 0.16 standard deviations on geometry subtest 2. Reasoning Mind: null</td>
<td>Kindergarten through fifth grade students in 3 schools in an elementary charter school network in San Francisco</td>
<td>583 students</td>
<td>Student</td>
</tr>
<tr>
<td>Wijekumar, Meyer, and Lei (2012)</td>
<td>Education Technology Research and Development</td>
<td>ITSS (Intelligent Tutoring for Structure Strategy) program for reading and language</td>
<td>Positive effects on language</td>
<td>0.10 standard deviations on language and 0.49 standard deviations on main idea quality</td>
<td>Rural and suburban fourth grade classrooms</td>
<td>2,643 students in 60 rural and 71 suburban classrooms</td>
<td>Classroom</td>
</tr>
<tr>
<td>Wijekumar et al. (2014)</td>
<td>Journal of Research on Educational Effectiveness</td>
<td>ITSS program for reading and language</td>
<td>Positive effects on literacy and signaling tests</td>
<td>0.20 standard deviations on literacy and 0.42 standard deviations on signaling tests</td>
<td>Fifth-grade classrooms in 45 schools within 12 school districts in rural and suburban settings in Pennsylvania</td>
<td>2,645 students in 58 rural and 70 suburban classrooms</td>
<td>Classroom</td>
</tr>
</tbody>
</table>
programs in math, considering models from light-touch homework supplements to class curriculum changes to school-wide personalized learning models. After reviewing the experimental evidence on math software, we then turn to the few experimental studies on CAL reading programs. Finally, we consider findings from the studies we included within the broader research context and highlight potentially promising directions moving forward. For our discussion, we draw our conclusions from well-powered evaluations that estimate the impacts of CAL programs with a high degree of precision.

Beginning with light-touch approaches, ASSISTments represents an especially promising example of a CAL program. ASSISTments is a math homework platform released by the Worcester Polytechnic Institute that does not require that schools adjust their curriculum or textbooks, and is available free of charge (ASSISTments 2016). The program is designed to carry out “formative assessments,” that is, to use “data from students’ independent work to give them helpful feedback and guidance while enabling the teacher to use the data to adjust instruction to meet students’ learning needs” (Roschelle et al. 2016). As students work through individual problems, the computer informs them about whether their answer is correct and offers guidance if necessary. Students are expected to benefit from the customized practice, as well as from rapid feedback and data supplied to teachers (in addition to, in some cases, supplementary professional development to train the teachers on optimizing use of ASSISTments). Referring back to our model, the ASSISTments program aims to support cognitive skill formation by providing students with immediate feedback on homework performance, a critical step in advancing one’s understanding of material and acquisition of content knowledge. Moreover, the program attempts to mitigate constraints a teacher typically faces when providing feedback and gathering timely information on student homework performance. Two small-scale proof-of-concept studies (Kelly et al. 2013, Singh et al. 2011) found promising effects, but these studies had samples numbering only in the dozens of students and implementation time numbering only in the days.

More recently, however, a full-scale impact evaluation of an ASSISTments intervention was conducted with a sample of 2,850 seventh graders across 43 schools in Maine (Roschelle et al. 2016). The authors found that the program improved math scores for treatment students by 0.18 standard deviations. This impact is particularly noteworthy given that treatment students used the program, on average, for less than ten minutes per night, three to four nights per week (Roschelle et al. 2016, p. 6). Students at or below median benefited even more from using the program, with an effect size of 0.29 standard deviations (Roschelle et al. 2016, p. 8). It is worth noting that the program depends on students’ ability to access a laptop or tablet—part of the reason that this evaluation was conducted in Maine, given its policy of lending laptops to all public middle and high school students. This further reinforces the promise of initiatives that aim to bolster the learning environment through increased access to technology while simultaneously supporting cognitive skill formation through educational software. While this hurdle may raise some external validity concerns with regard to this particular study, a variety of possibilities exist for enabling access in other states, especially given that software and licensing are free, so costs are otherwise low.

Other programs move beyond homework supplements and instead offer full curricula. A prime example—perhaps the most prominent of all of the CAL products discussed in this review—is the set of Cognitive Tutor programs published by Carnegie Learning.
Unlike ASSISTments, the Cognitive Tutor programs generally provide curricula for entire mathematics courses, including lesson plans, textbooks, training for teachers, and detailed guidelines. When viewed in the context of our framework, Cognitive Tutor may support teachers in promoting cognitive skills formation by reducing time to develop lesson plans while simultaneously delivering more targeted content. The company recommends 40 percent computer time and 60 percent class time (Cabalo, Ma, and Jaciw 2007; Pane et al. 2010). Through the “tutor,” students receive individualized instruction in the form of challenging problems that reflect real-world situations, enabling students to move from concrete to abstract thinking (Pane et al. 2014). We identified nine papers reporting on experimental studies on Cognitive Tutor programs in a variety of locations, including California, Hawaii, Maryland, and Oklahoma (Cabalo, Ma, and Jaciw 2007; Campuzano et al. 2009; Dynarski et al. 2007; Karam et al. 2017; Morgan and Ritter 2002; Pane et al. 2010; Pane et al. 2014; Ragosta, Holland, and Jamison 1982; Ritter et al. 2007). Earlier papers were narrow in scope and had mixed results ranging from negative effects of 0.19 standard deviations (SDs) to positive effects of 0.23 SDs on math outcomes. One 2010 experiment of Cognitive Tutor’s geometry program found that students who used the software performed worse than students who were taught the standard geometry curriculum. Qualitative analysis from the study suggests that teachers had difficulty implementing the program in a way that engaged students, potentially because the program represented too large of a shift from the business-as-usual approach. In particular, the study noted that teachers struggled to effectively facilitate group work required by the curriculum and also often struggled to connect students’ individualized work in the self-paced modules to what was being taught in class (Pane et al. 2010). Rather than reducing constraints on teacher time, the introduction of the software may have been an additional constraint.

A 2014 experiment across eight states in 73 high schools and 74 middle schools has sought to increase the external validity of the Cognitive Tutor literature by seeking to replicate realistic scale-up conditions in a wide variety of locations (Pane et al. 2014). They found no effect the first year, but a 0.20 standard deviation impact with the second-year cohort. Interestingly, the improvement in the second-year cohort was not associated with increased fidelity of implementation, but instead with teachers reducing (although not completely eliminating) their use of the activities called for by Cognitive Tutor guidelines for noncomputer class time (Pane et al. 2014).

While many studies have posited that quality of implementation is an important factor in determining outcomes (Pane et al. 2010, Pane et al. 2014, Ritter et al. 2007), one study examining implementation fidelity of the Cognitive Tutor is more specific about how quality of implementation relates to time use in the classroom. Adaptive software may ensure that the amount of time students spend engaged on the platform is time spent learning academic concepts, rather than playing games or partaking in other nonacademic tasks. Their findings suggest that learner efficiency is associated with whether teachers take an active role in turning the time students are engaged in activities required by the software into academic learning (Fancsali et al. 2016). Such variations in implementation may explain mixed results across the Cognitive Tutor studies, and at least warrant further review for understanding the extent to which this theory about time use describes a true mechanism of impact. Moreover, if one goal of CAL is to help teachers overcome constraints in the teaching process, it raises important open questions for how teachers should be deploying software to best do so.
Another intervention that has recently risen to prominence is SimCalc. Although SimCalc has not been used or tested as extensively as Cognitive Tutor programs, those studies that have been conducted demonstrate strong potential. SimCalc aims to improve the instruction of mathematics relating to algebra and leading to calculus (Roschelle et al. 2010) by supporting cognitive skill formation through computer-assisted modeling techniques. Using methods of “representational infrastructure,” the program enables students to control the motions of animated characters by building or editing mathematical functions. After editing the functions, students can press a “play” button to view the corresponding animation (Kaput and Roschelle 2013). A study on a SimCalc intervention in 140 seventh grade and 80 eighth grade Texas classrooms turned up one of the largest effect sizes of any large-sample study covered in this review, with 0.63 and 0.56 standard deviation improvements in math scores for seventh and eighth graders, respectively (Roschelle et al. 2010). Although it is unknown whether these effect sizes would translate to other contexts outside of the Texas public school system, the study did attempt to recruit a diverse sample of teachers and students, suggesting its relevance to varying school populations. Any single effect size, and particularly one of this magnitude, must be taken with caution—but an effect of even half this size would be noteworthy.

We identified only seven studies (Borman, Benson, and Overman 2009; Deault, Savage, and Abrami 2009; Faber and Visscher 2018; Mitchell and Fox 2001; Rouse and Krueger 2004; Wijekumar, Meyer, and Lei 2012; Wijekumar et al. 2014) within the developed world exclusively examining reading and spelling programs. Of these, two evaluated Fast ForWord, a program initially designed for students with particular learning disabilities (Rouse and Krueger 2004; Borman, Benson, and Overman 2009), but that has been, in some cases, marketed and used to cope with broader reading challenges. The program works by providing students with individualized exercises in a game-like computerized environment, where students receive on-screen rewards for correct answers and attentiveness to instruction. Lower-performing students were pulled out of the classroom to devote 90–100 minutes of time to the software five days a week for six to eight weeks, so the software acted as a substitute for other classroom curriculum. These studies—the only ones, to our knowledge, that have evaluated Fast ForWord within a broader education setting—found mostly weak and insignificant results, although one study only had a sample of 89 students. While the specific mechanisms were not explicitly tested, the study did note that the supervisor overseeing the computer sessions needed to know how to proactively identify and help students who were struggling (Rouse and Krueger 2004). Hence, it seems that the software did not utilize adaptivity or the fast feedback of data features that other effective CAL programs retain. In other words, it may not have helped instructors overcome time and resource constraints to the degree that effective CAL programs did. While small effect sizes of the Fast ForWord program cannot be entirely ruled out, experimental evidence suggests that further adjustments or at least more testing may be needed before scale-up can be recommended.

In contrast, two recent well-powered studies (Wijekumar, Meyer, and Lei 2012; Wijekumar et al. 2014) that evaluated a reading comprehension program called Intelligent Tutoring for the Structure Strategy (ITSS), which teaches students a particular technique for breaking down texts, show significant positive results. It differs from Fast ForWord in that it is geared toward middle school students and aims to improve reading comprehension rather than basic...
literacy. ITSS is a web-based intelligent tutor that utilizes a “structure strategy” to teaching literacy that begins a lesson by describing what the student is going to learn, models the strategy, and asks the student to practice. The tutor then provides feedback to the student based on his/her answers and gives the student the chance to correct the answer if needed. In this way the program seeks to mitigate the aforementioned constraints on teacher time by replicating how an in-person tutor might help a student systematically break down a passage of text. Effect sizes on a series of reading comprehension measures ranged from 0.2 to 0.53 standard deviations. Though further research is needed to better tease out the underlying mechanisms and ascertain external validity beyond the schools recruited for this study, the use of the “structure strategy,” which simulates for students how effective readers process text, may be a key driver of impact, as opposed to a program like Fast ForWord, which works more through reward and practice. Moreover, ITSS offered students immediate feedback on answers and the opportunity to correct answers if needed, which again suggests that personalization may be an important underlying mechanism to the successful formation of cognitive skills.

CAL is becoming increasingly popular within the developing world as well, and an experimental literature on these interventions is growing rapidly in China (Bai et al. 2016; Lai et al. 2013; Lai et al. 2015; Lai et al. 2016; Mo, Zhang, Luo, et al. 2014; Mo, Zhang, Wang, et al. 2014; Mo et al. 2015) and India (Banerjee et al. 2007; He, Linden, and MacLeod 2007; Linden 2008; Muralidharan, Singh, and Ganimian 2016; Naik et al. 2016). On one hand, CAL programs may prove to be more effective in developing countries given the often tight capacity constraints faced, such as low supply of qualified teachers or large pupil-to-teacher ratios. On the other hand, infrastructure limitations and other challenges could impede CAL implementation. Findings so far have been largely positive.

One recent study conducted in Delhi (Muralidharan, Singh, and Ganimian 2016) finds especially large effects that seem to occur through mechanisms of personalization and adaptivity akin to those described above. The program, called Mindspark, administers its self-developed educational software at study centers for a small fee. The Mindspark program conducts an initial assessment to determine a student’s learning level and then continually adapts content to match a student’s learning level based on how the student progresses through earlier content. After a treatment period of under five months, the authors find an effect of 0.36 standard deviations on math scores and 0.22 standard deviations on Hindi language scores, the two subject areas for which the program was tested. Although there is no treatment arm that offers the same content without the adaptivity component, they present strong suggestive evidence that adaptivity played a key role in accounting for the impact. There is an expansive range of levels between students within each grade, and the Mindspark program records that report the questions generated show that they matched this wide range. Given that no teacher could possibly have covered such a huge spread of levels, the authors argue that the adaptation element of the program must have played a central role in enabling its positive impact and could therefore be an integral part of a solution to the unevenness of levels that challenge many schools in India and elsewhere. From an implementation standpoint, CAL may be useful in solving gaps in staffing or infrastructure problems, which can be particularly useful in developing country contexts where there is low supply of qualified teachers, or teachers are less likely to be computer literate. However, it could also be useful in US contexts where the binding constraints...
to learning are similar, for example in rural areas to supplement low quality or low supply of teachers.

Taken together, CAL may be able to significantly improve learning outcomes, with the evidence particularly strong for math. By using technology to provide students with rapid feedback and teachers with data on student performance, supplementary programs can overcome long-standing classroom constraints and yield significant benefits to cognitive skill formation (Roschelle et al. 2016). Promising studies of more intensive interventions also suggest that technology has the potential to replicate aspects of in-person tutoring for both math and reading (Roschelle et al. 2010, Wijekumar et al. 2014).

Nevertheless, there are some caveats to note when drawing inferences from the results of these studies. First, it may very well be that the promising CAL interventions are the ones that are more likely rigorously evaluated. Hence, it is important to consider the potential mechanisms of impact for each intervention and not overgeneralize to infer that the results from these 31 studies imply that all CAL interventions are promising. Rather, we should consider which specific features of CAL interventions have been shown to be efficient in improving returns to investment in cognitive skill formation, such as personalized learning, adaptivity, and rapid feedback. Assuming teacher quality is a critical input into the education production function, identifying CAL programs that successfully enhance teachers’ ability to deliver instruction at scale could yield profound benefits.

4.3 Looking Forward

Numerous important questions remain, however, for future researchers to answer if CAL’s potential is to be efficiently leveraged. One vital area is to more explicitly test for the underlying mechanisms—whether they are adaptivity of instruction, rapid feedback to teachers, or more engaging delivery of content—that most contribute to the success of effective programs.

Another crucial question is to test the extent to which learning from CAL lasts in the longer term. To what extent do effects compound or diminish in subsequent years? Another important task will be to further explore whether and when CAL can work effectively for subjects other than math. Do the cognitive processes that underpin mathematical reasoning inherently lend themselves better to software algorithms? More broadly, which areas of education could CAL add most value to, and which pedagogical strategies are best enhanced by CAL? And when are light- versus heavy-touch interventions most appropriate and cost effective? An important crosscurrent that undercuts many of these other concerns is the issue of implementation. One way to gain greater leverage on this issue could be to test a particular CAL program in a particular population while varying elements of the implementation plan. Finally, we still know little about how CAL programs interact with teachers’ efforts. Unpacking interconnections could highlight opportunities for how CAL can best complement and increase the efficiency of teacher’s efforts. As technology continues to evolve, testing new emerging CAL models will be vital.

5. Online Courses

Since their emergence during the 1990s, online courses have come to constitute a sizable presence within the education field. By 2013, over a third of US college students had taken an online course at some point during their college career (Bettinger et al. 2017, citing Allen and Seaman 2013), and more than 11 percent were enrolled in entirely online programs (Deming et al. 2016). The rise of online learning bears heavily on policy issues
relating to educational equity, since two key justifications for the proliferation of online education have been its promise of improving access and reducing marginal costs associated with teaching more students. Moreover, at least at the postsecondary level, students in online programs tend to have faced significant disadvantages. For instance, data from the National Postsecondary Student Aid Study's 2010/2011 representative survey indicate that “online students are older, have lower levels of parental education, are more likely to be single parents themselves, and are more likely to be working full-time while enrolled in school than are other college students” (Deming et al. 2016). So how does online education perform in terms of access, learning, and other important outcomes? How does online education impact the learning environment and the formation of cognitive skills?

Online courses have, over the past several years, coalesced into two broad categories. First, what we refer to as conventional online courses represent an online extension of the “distance learning” or “correspondence course” format, an approach that has a long history in higher education (Means et al. 2009). These courses are typically offered as part of a degree program that consists entirely of online courses, or that includes online, face-to-face, or blended⁹ courses. Second are massive open online courses (MOOCs). Unlike conventional online courses, MOOCs are typically not part of official degree programs and are open to anyone to enroll via the internet. They broadly consist of “structured and sequenced teacher-led activities (e.g., videos, readings, problem sets) coupled with online assessments and usually some venue for student interactions such as a discussion forum” (Hodges, Lowenthal, and Grant 2016). Between 2012 and 2015, MOOCs saw enrollment rates exceeding 25 million (Kizilcec et al. 2017). While conventional online courses and MOOCs developed to serve largely separate purposes, the lines between them are becoming blurred. For instance, MOOC companies have increasingly offered certification programs for a fee, such as MicroMasters programs,¹⁰ and MIT has launched a MOOC-based program that will lead to a traditional master’s degree.¹¹

In this section, we first discuss the background, context, potential mechanisms of impact, and the experimental evidence on conventional online courses, and then turn to a discussion of the same with regard to studies on MOOCs.

5.1 Conventional Online Courses

5.1.1 Background, Context, and Mechanisms of Impact

Online courses build on a tradition of correspondence courses that has existed for over a century within the higher education field (Means et al. 2009). As early as the late 1800s, institutions like the University of Chicago and the University of Wisconsin were teaching faraway students via the postal service (Deming, Goldin, and Katz 2012). Educators and entrepreneurs brought online college courses and degree programs to market beginning in the 1990s, but proliferation expanded rapidly after a 2006 decision to end a regulation that had limited federal aid money for institutions conducting more than half of their coursework via correspondence (Deming et al. 2016). Some institutions offer both online and face-to-face instruction,

⁹The term blended takes on different meanings in different contexts within the ed-tech literature—in this case, we use the term to refer to a single course that has both online and face-to-face components.

while others offer online courses exclusively. While a growing mass of selective universities offers online programs, large, for-profit colleges (Deming, Goldin, and Katz 2012) like the University of Phoenix (McKenzie 2018) continue to control a large share of the market, although nonprofit online universities have experienced a recent surge in enrollment.

How, then, might conventional online courses add value to education? Thus far, experimental research on conventional online courses has compared online against face-to-face courses to judge the extent to which the former improves access and can act as a viable substitute for face-to-face education. In the framework presented in this review, their aim is to identify whether online learning cultivates an improved learning environment—by expanding access to education—and improves test scores or academic performance—signals of improvement in the formation of cognitive skills—at a given level of cost per student.

However, online courses may work through different channels to achieve impacts on learning. For example, a key justification for online courses is that, in many contexts, they may be much less expensive to implement than face-to-face courses. “If Internet-based classes are at least reasonable substitutes for live lecture classes, then the use of Internet-based classes could be a cost-effective method of combating increased fiscal constraints” (Figlio, Rush, and Yin 2013; see also Cowen and Tabarrok 2014, Means et al. 2009). Second, online courses can expand access by allowing people to take courses or entire degree programs that would not otherwise be possible or worthwhile for them to take, for instance, because of geographic location, work or family obligations during class hours, or disabilities (Goodman, Melkers, and Pallais 2019; Means et al. 2009; Poirier and Feldman 2004). And online courses may allow students more flexibility in accessing course materials at the most convenient times, and in spending more time on content that they are struggling with and less on content that they have mastered (Figlio, Rush, and Yin 2013). This potential for online courses to increase access and lower cost may, in some cases, constitute outcomes in their own right, and in some cases these may be necessary conditions for improving academic performance and the formation of cognitive skills.

Educators and researchers have also pointed out potential drawbacks of online courses. The flip side of online courses’ flexibility is that students who do better with externally induced structure may be more likely to face time management issues than they would for a face-to-face class, and may thus fall behind (Donovan, Figlio, and Rush 2006). It is also possible that too large a shift toward online courses could take away opportunities for networking and interaction that arise more naturally in face-to-face environments (Sleeter 2014). More generally, some educators and researchers believe that a valuable element of the teaching process is lost when the face-to-face dimension is reduced or eliminated (Sleeter 2014).

5.1.2 Experimental Evidence

We identified 17 experimental studies examining the effects of conventional online courses. Of these, 15 RCTs compared...
online and face-to-face delivery (or various gradations in between) of particular courses. Four of the 15 RCTs (Esperanza, Fabian, and Toto 2016; Foldnes 2016; Harrington et al. 2015; Wozny, Balser, and Ives 2018) examine the impact of flipped classroom models, a type of blended learning that reverses the traditional learning environment by delivering instructional content, which is traditionally taught inside the classroom, at home via the internet (Tucker 2012). One RDD (Goodman, Melkers, and Pallais 2019) tested the extent to which offering an online degree option increased enrollment, and one audit RCT tested whether employers distinguished between online and face-to-face degree when selecting resumes to follow up on (Deming et al. 2016).

To what extent does the evidence suggest that online classes can match or exceed learning outcomes from face-to-face classes? While a great deal more exploration and replication would be needed to draw robust conclusions, the studies reviewed here are consistent with the hypothesis that, without some degree of face-to-face teaching, learning outcomes may suffer, leading to (albeit small) sacrifices in test scores for fully online courses relative to face-to-face courses. In contrast, blended learning environments—meaning, in this case, courses that have both a face-to-face component and an online component—have not been found to significantly underperform purely face-to-face courses in studies meeting our methodological criteria. So, while the evidence at this point would not back substantial shifts toward fully online courses, it does indicate that switching courses from fully in-person to blended could decrease costs without negatively affecting quality.

The first full-scale field experiment to compare face-to-face with online courses took place in an introductory economics course in 2013 at a major research university with a sample of over 300 students.\footnote{Keefe (2003) conducted a related study in an undergraduate business course and comes up with results that are in the same direction as Figlio, Rush, and Yin (2013), but this study had a sample of only 35 students (with students in face-to-face classes performing better). Another study conducted in a university psychology class with a sample of only 23 students found opposite results, with students in the online version performing marginally better than those in a face-to-face group. Zhang (2005) and Zhang et al. (2006) run experiments on 155 and 138 undergraduates respectively and find that interactive online modules outperform noninteractive online modules and face-to-face sessions, but the context is single-session lab experiments rather than a field experiment with actual classes.}

The course was identical for all students, but some students were provided access to online video lectures, while others attended these lectures in person. The study finds that students in the in-person group show higher outcomes, and that the differences are relatively small but significant—around 3 percentage points on the midterm and about 2.5 percentage points on the final. In actual university settings however, the choice will not necessarily be between courses that are entirely face-to-face or entirely online. Instead, the two are often mixed into blended courses. Two subsequent experiments studied blending learning environments of this sort. One compared outcomes for a statistics course in which one group received three hours per week of face-to-face instruction time, while another group received only one hour of instruction time but additional internet-based exercises. The second experiment tested the effects of reducing face-to-face time in an economics course where all students also had access to online resources. Neither experiment found significantly better outcomes to be associated

(under the Computer Assisted Learning section), experimentally evaluates a middle school summer math program that includes an hour daily use of Khan Academy, but since the study compares the program as a complete package against a control group that does not attend any program, the study cannot identify independent effects of the online component.
with more in-person class time in a blended learning context.

Finally, the most comprehensive study in this strand of the literature—the only one to test fully online, blended, and fully face-to-face courses within the same experiment—found results consistent with each of the above (Alpert, Couch, and Harmon 2016). Here, the authors test the impact in an economics course of two treatment arms—one purely online and one blended—along with a fully face-to-face control group in a single experimental context. This study finds that students in the purely online version of the course do not perform as well as those in the purely face-to-face group, while outcomes for the blended treatment group are not statistically different from the control (Alpert, Couch, and Harmon 2016).

The majority of research on online courses has been conducted in postsecondary settings, but educators have increasingly attempted to leverage online learning in middle and high school environments as well. We identified two such experimental studies, both of which tested the effect of offering an online version of algebra content. One study tested the effectiveness of online summer credit recovery courses relative to face-to-face courses for students who had failed freshman algebra (Heppen et al. 2012). This study was conducted in 15 high schools in the Chicago Public Schools system with the lowest rates of students passing freshman algebra, with a sample of nearly 1,400 students across two cohorts. The hope was that the online course would provide “a more individualized, interactive experience” prompting students to “be more engaged and more likely to persist in the course.” However, students in the face-to-face course outperformed those in the online course. Suggestive evidence from the study indicates that one significant reason was that teachers in the face-to-face course were better able to flexibly incorporate a range of topics, and thus were better able to accommodate and engage the students (Heppen et al. 2012).

A complementary study conducted in schools in Maine and Vermont evaluated the effect of offering online Algebra I content to grade 8 students who were ready to take algebra, but who attended schools that did not typically offer a full-section of Algebra I in grade 8 (Heppen et al. 2011). In the study, treatment schools offered a full online Algebra I course to eligible students, while control schools offered the regular math curriculum, which included some Algebra I, but not the full course. The study found that taking the online algebra course significantly improved students’ algebra achievement at the end of grade 8 and increased the likelihood of participating in advanced course-taking sequences in high school. Here, the mechanism is similar to other studies of online courses in which the online component facilitates access to content students would otherwise not be able to access, which in turn potentially improves the formation of cognitive skills. Nevertheless, while this study yielded positive results, the majority of the schools that agreed to participate in the study were small rural schools. Hence, the generalizability of the findings to schools outside this sample is unknown and further research to understand longer-term effects of similar online course offerings is needed.

While these studies tell us something about how online learning may increase access to discrete curricula or content, the next logical question is: to what extent can online courses increase access to further education for those who face barriers to pursuing a face-to-face degree at all? One of the main justifications for the potential usefulness of online courses is that they can improve access to degree programs for populations that otherwise might have trouble accessing them. We identified only a single study fitting our criteria that addressed this
question. Specifically, the researchers relied on an RDD design to reveal that prospective students applying to Georgia Institute of Technology’s online master’s program in computer science who were just above an admissions cutoff (which was not known to the applicants) for the online version of the program were 20 percentage points more likely to eventually enroll in the online program than those just below the cutoff (Goodman, Melkers, and Pallais 2019). Given that the estimates also showed a 20 percentage point difference in the probability of admission to the online program between those just above and below the cutoff, the results indicate that almost all of the additional admits actually attended the college. The strongest effects were observed among mid-career prospective students, who otherwise may have chosen not to complete a degree at all had the online program not been offered to them (Goodman, Melkers, and Pallais 2019). Another recent experiment, however, finds that “a business bachelor’s degree from a for-profit online institution is 22 percent less likely to receive a callback than one from a non-selective public institution” (Deming et al. 2016). But the design does not allow for disentangling the effect of the education medium (online versus face-to-face) from the institution’s for-profit/not-for-profit status. And even if employers do place a penalty on online degrees, this may change in the coming years given the ongoing expansion of the online education sector.

5.1.3 Looking Forward

Overall, the experimental research on conventional online courses has explored their potential to reduce costs, strengthen the learning environment, and facilitate the formation of cognitive skills through increasing access to discrete curricula or entire degree programs. The evidence so far indicates a potential for blended learning to reduce costs without lowering quality; however, there are some cautions for moving toward a purely online curriculum given the evidence of null or negative effects on test scores. While the specific mechanisms are unclear, it may be that less structure for time management or less opportunity to learn from peers or instructors in a purely online setting is detrimental to learning. Nevertheless, we do see that online instruction can, in some cases, improve the learning environment by increasing access when it is otherwise not possible, which can in turn facilitate the formation of cognitive skills. However, the evidence remains thin and mixed, suggesting that further research is still needed to disentangle the potential mechanisms. For example, are the null or negative effects of pure online learning due to a loss of personal interaction with teachers or peers, less structured processes to help students with time management, an inability for instructors to adjust lessons to the students’ particular needs, or some combination of the above? If this is the case, online courses may have difficulty supporting the formation of certain noncognitive skills, such as perseverance, for which in-person instruction may be better designed. Additionally, the nature of study recruitment and participation, and the fact that most studies have been conducted with introductory economics, statistics, or math courses may pose concerns with external validity. It may be that online learning may be relatively more or less beneficial to these particular academic subjects, or the study populations involved. Further studies to test online learning in other subjects with other populations and specifically designed to disentangle these mechanisms are needed to fully understand the potential of conventional online learning courses, and to better understand potential design solutions.

With these lessons from conventional online learning studies in mind, we now turn to a discussion of the experimental evidence
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpert, Couch, and Harmon (2016)</td>
<td><em>American Economic Review</em></td>
<td>Face-to-face versus blended versus purely online course content</td>
<td>1. Negative effect on learning for purely online versus face-to-face course 2. Null effect on learning for blended versus face-to-face course</td>
<td>Main reported effect sizes (not standardized): 1. 4.9 points lower in purely online course relative to face-to-face course 2. Null</td>
<td>College students of a principles of microeconomics course taught at a large public university in the Northeast.</td>
<td>519 students</td>
<td>Student</td>
</tr>
<tr>
<td>Bowen et al. (2014)</td>
<td><em>Journal of Policy Analysis and Management</em></td>
<td>Blended instruction versus face-to-face only</td>
<td>Null effect on learning for blended versus face-to-face course</td>
<td>Null</td>
<td>6 public university campuses in the United States</td>
<td>605 students</td>
<td>Student</td>
</tr>
<tr>
<td>Deming et al. (2016)</td>
<td><em>American Economic Review</em></td>
<td>Audit of fictitious resumes varied by for-profit versus public, online versus brick-and-mortar, and more selective versus nonselective postsecondary institutions, based on degrees and programs in business and health</td>
<td>Negative effect on callback rates for business bachelor's degrees from for-profit online institution</td>
<td>Main reported effect sizes (not standardized): 22 percent decrease in likelihood of callback relative to a nonselective public institution.</td>
<td>Employers posting job vacancies in business and health identified by a nationally recognized online job search website in Chicago, Los Angeles, Miami, New York City, and San Francisco</td>
<td>10,492 resumes</td>
<td>Resume</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Esperanza, Fabian, and Toto (2016)</td>
<td><em>Adaptive and Adaptable Learning</em></td>
<td>Flipped classroom model</td>
<td>Positive effect on enjoyment in math; positive effect on math scores (under analysis of covariance (ANCOVA) specification)</td>
<td>1. 0.31 standard deviation increase in reported enjoyment in math; 2. 0.59 standard deviations ( (p = .09) ) increase on math scores; significant under ANCOVA specification</td>
<td>High school algebra students</td>
<td>91 students</td>
<td>Student</td>
</tr>
<tr>
<td>Figlio, Rush, and Yin (2013)</td>
<td><em>Journal of Labor Economics</em></td>
<td>Online lectures</td>
<td>Negative effect on learning outcomes</td>
<td>Main reported effect sizes (not standardized): 2.508 (out of 100) point decrease relative to live courses.</td>
<td>University students enrolled in a large principles of microeconomics course</td>
<td>327 students</td>
<td>Student</td>
</tr>
<tr>
<td>Foldnes (2016)</td>
<td><em>Active Learning in Higher Education</em></td>
<td>Flipped classroom model</td>
<td>Positive effect on math scores</td>
<td>Main reported effect sizes (not standardized): 12 percentage point increase in performance relative to the lecture group</td>
<td>First-year undergraduate business students</td>
<td>235 students</td>
<td>Student</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goodman, Melkers, and Pallais (2019)</td>
<td><em>Journal of Labor Economics</em></td>
<td>Online Master of Science in Computer Science</td>
<td>Positive effect on overall enrollment in formal education</td>
<td>Main reported effect sizes (not standardized): 20 percentage point increase in overall enrollment in formal education; at least 7 percent increase in annual production of American computer science master’s degrees</td>
<td>Online and in-person applicant pools for Georgia Institute of Technology’s online and in-person computer science master’s program</td>
<td>17,632 students</td>
<td>Not randomized—regression discontinuity design</td>
</tr>
<tr>
<td>Harrington et al. (2015)</td>
<td><em>Nursing Education Perspectives</em></td>
<td>Flipped classroom model</td>
<td>Null effect on grades and exam scores.</td>
<td>Null</td>
<td>Nursing students</td>
<td>82 students</td>
<td>Student</td>
</tr>
<tr>
<td>Heppen et al. (2011)</td>
<td><em>Institute of Education Sciences</em></td>
<td>Online Algebra I course (for students who would otherwise have no access to algebra)</td>
<td>Positive effects on eighth grade Algebra I achievement and likelihood of participation in advanced math course sequence in high school</td>
<td>1. 0.40 standard deviation increase in eighth grade algebra achievement  2. 1.96 times as likely to do advanced math course sequence in high school</td>
<td>Schools in Maine and Vermont that serve students in grade 8 and did not offer a stand-alone algebra class</td>
<td>445 students</td>
<td>School</td>
</tr>
<tr>
<td>Heppen et al. (2012)</td>
<td><em>Society for Research on Educational Effectiveness (SREEd) Conference Paper</em></td>
<td>Online algebra courses for credit recovery</td>
<td>1. Negative effects on grades for online students versus students in the face-to-face course  2. Negative effects on credit recovery</td>
<td>1. 0.18 standard deviation decrease in end-of-course assessment score  2. 7 percentage point decrease in credit recovery</td>
<td>2 cohorts of students at Chicago Public Schools who failed Algebra I in ninth grade and enrolled in summer recovery program.</td>
<td>592 students</td>
<td>Student</td>
</tr>
</tbody>
</table>
### TABLE 3A
TRADITIONAL ONLINE COURSES (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jackson and Makarin (2018)</td>
<td><em>American Economic Journal: Economic Policy</em></td>
<td>Teacher access to online off-the-shelf quality lessons and support to promote their use</td>
<td>Positive effect on math achievement</td>
<td>1. 0.06 standard deviations from access alone 2. 0.09 standard deviations from access and support</td>
<td>All middle school teachers in 3 school districts in Virginia</td>
<td>27,613 students in 363 classrooms</td>
<td>Classroom</td>
</tr>
<tr>
<td>Joyce et al. (2015)</td>
<td><em>Economics of Education Review</em></td>
<td>1 class/week (blended) versus 2 classes/week (face-to-face)</td>
<td>Positive effect on midterm score for students in face-to-face format; null effect on final score.</td>
<td>0.21 standard deviation increase on midterm scores; null on the final</td>
<td>Students at Baruch College in microeconomics course</td>
<td>725 students</td>
<td>Student</td>
</tr>
<tr>
<td>Keefe (2003)</td>
<td><em>Educause Quarterly</em></td>
<td>2 studies: 1. Lecture and interaction online versus traditional face-to-face 2. Interaction versus regular lecture experience</td>
<td>Main reported effect sizes (not standardized): 1. 0.7 decrease on instructor satisfaction and 0.6 decrease (on scale out of 6.0) relative to face-to-face course 2. 7.6 percent decrease in exam scores relative to the face-to-face course</td>
<td></td>
<td>6 sections of students in an organizational behavior course in Indiana University Southeast</td>
<td>118 students</td>
<td>Student</td>
</tr>
<tr>
<td>Poirier and Feldman (2004)</td>
<td><em>Teaching of Psychology</em></td>
<td>Traditional face-to-face versus online course</td>
<td>Positive effect on grades for students in the online course.</td>
<td>1. 0.57 standard deviation increase in overall course exam 2. Null on paper assignments compared to students in the traditional course</td>
<td>Students from a large state university who indicate that either a face-to-face or an online course was acceptable</td>
<td>23 students</td>
<td>Student</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wozny, Balser, and Ives (2018)</td>
<td><em>Journal of Economic Education</em></td>
<td>Flipped classroom model</td>
<td>Positive effect on high-stakes assessments</td>
<td>0.16 standard deviation increase relative to a lecture class</td>
<td>Introductory undergraduate econometrics course at the United States Air Force Academy</td>
<td>137 students</td>
<td>Student</td>
</tr>
<tr>
<td>Zhang (2005)</td>
<td><em>American Journal of Distance Education</em></td>
<td>The interactive e-classroom component of the Learning By Asking (LBA) system versus traditional face-to-face classrooms</td>
<td>Positive effect on performance and satisfaction for students in the fully interactive multimedia based e-learning environment.</td>
<td>1.18 standard deviations from fully interactive LBA system; 0.45 standard deviations from less interactive LBA system.</td>
<td>Undergraduate students from a large public university in the United States</td>
<td>155 students</td>
<td>Student</td>
</tr>
<tr>
<td>Zhang et al. (2006)</td>
<td><em>Information &amp; Management</em></td>
<td>Non-interactive video learning environments</td>
<td>Null effects for students who used the e-learning environment that provided noninteractive video.</td>
<td>Positive impacts (see entry above); null for e-learning environment that provided noninteractive video</td>
<td>Undergraduate students from a large university in Southwest United States</td>
<td>138 students</td>
<td>Student</td>
</tr>
</tbody>
</table>
on MOOCs before discussing future directions for both types of online learning.

5.2. Massive Open Online Courses (MOOCs)

5.2.1 Background, Context and Mechanisms of Impact

The term MOOC was first used in 2008 by media theorists George Siemens and Stephen Downes for a course they taught at the University of Manitoba entitled “Connectivism and Connected Knowledge,” with 25 students participating in face-to-face sessions at the university and content broadcasted to 2,300 additional students via the internet (Greene et al. 2015, see also Cormier and Siemens 2010). In the subsequent decade, MOOCs have proliferated rapidly, with thousands of courses offered and hundreds of thousands of students enrolled worldwide (Greene et al. 2015). Like online courses, educators and education policy makers saw in MOOCs the potential to decrease costs and increase access (Greene et al. 2015). Because MOOCs are generally “open,” they have the potential to reach exponentially more students in a much more diverse range of contexts than can conventional online courses granted for credit. However, because MOOCs usually do not build toward a degree and may or may not be valued on the labor market, it is less clear what, if any benefits, MOOCs may bring beyond the value of the educational content they impart.

What has been the effect of MOOC proliferation? Observational research has found that expectations that MOOCs will “democratize education” have been overblown and that, although MOOCs have offered the opportunity for many disadvantaged individuals to access high-quality educational content, enrollment and success rates are highly skewed toward advantaged populations. Thus, MOOCs may even “exacerbate rather than reduce disparities in educational outcomes related to socioeconomic status” (Hansen and Reich 2015). But their overall impact on learning outcomes is difficult to evaluate. While researchers are interested in the effects of MOOCs on education, it is less clear what to compare them to since they generally do not substitute for face-to-face courses that students would otherwise take. People may take MOOCs for a wide variety of reasons, from practicing skills for school or work to fun and personal interest. Because MOOCs, broadly speaking, lack a clear counterfactual in that there is no single function they seek to fulfill or institution they attempt to substitute for, no clear experimental evidence has yet emerged on their overall impact, although this is likely to change over the next several years given the outpouring of interest. Nonetheless, MOOCs are being given to millions of students each year (Shah 2018), and researchers have begun to delve experimentally into questions of how MOOC usage can be improved for interested students. In fact, MOOCs lend themselves well to low-cost RCTs, among other types of data generation and analysis (Lamb et al. 2015).

Hence, in the framework presented in this study, the potential paths toward increased cognitive skills tested in the experimental research on MOOCs are slightly different from those expected in conventional online courses. While, as for online courses, the goal of MOOCs is ultimately to improve academic and other cognitive skills, experiments associated with MOOCs have focused on evaluating strategies to build or overcome gaps in noncognitive skills that prevent students from fully engaging with MOOCs. In particular, MOOCs face very

15We forego an in-depth discussion of ed-tech and noncognitive skills in this section since it constitutes a central element of the next section.
low completion rates: “few users actually complete the class” (Banerjee and Duflo 2014). These low rates in themselves do not necessarily signal a problem: many students enroll with no intention of completing the course, and students may generally be getting what they want from the MOOCs even if they are only accessing bits and pieces. But low rates may, at least in part, reflect missed learning opportunities that could be avoided with modifications to the MOOC platform (Banerjee and Duflo 2014). Thus, experimental research on MOOCs up to this point has focused primarily on whether and how a range of behavioral interventions can improve MOOC completion rates and extend coverage to disadvantaged groups, for example, by increasing interest, effort, and persistence.

Interventions aiming to improve student MOOC effort have generally followed the approaches of the behavioral and psychosocial interventions that will be discussed in greater detail in the following section. These interventions use strategies such as social comparisons, nudges to overcome procrastination or to increase participation in the online platform, or writing exercises to reduce social identity threat. In contrast to evaluations of conventional online courses, which primarily test the extent to which online courses improve cognitive skill formation, the experimental evidence on MOOCs primarily tests interventions that aim to change behavioral outcomes through psychological nudges. In this sense, these nudges are successful to the extent that they can help overcome a constraint to student engagement and persistence with the MOOC platforms which, in turn, is expected to lead to a low-cost boost in cognitive or academic skills.

5.2.2 Experimental Evidence

We identified 11 studies evaluating MOOC interventions. Seven of the 11 studies found positive effects from at least one treatment arm. Through what channels did these effects occur? One approach adopted from the behavioral economics literature has been the model of “social comparison” interventions—programs that inform students of their performance relative to other students. The behavioral economics literature suggests that social comparisons may drive individuals to try harder to excel. Two recent RCTs (Davis et al. 2017, Martinez 2014a) found that social comparison interventions can improve MOOC performance and completion, although one of these (Martinez 2014a) found significant effects only when framed “negatively” (i.e., when target students were informed of how many students had outperformed them rather than how many students they had outperformed).

Even if fully motivated to succeed in a course, MOOC students may struggle with time management issues and, in particular, the temptation to procrastinate. Procrastination may be a particularly acute temptation for MOOC students since they are not being directly observed by an instructor. One study that attempted to address problems of procrastination found that a commitment device that encouraged students to commit to limitations on time spent on distracting internet sites increased the likelihood of completion by 40 percent and grades by 0.29 standard deviations, while treatment arms that reminded students how much time they were spending on these websites or blocked them while on the course page showed no significant effect for both completion and academic performance (Patterson 2018). Banerjee and Duflo (2014), Davis et al. (2017), Kizilcec et al. (2014), Kizilcec et al. (2017), Lams et al. (2015), Martinez (2014a, b), Patterson (2018), Yeomans and Reich (2017). Banerjee and Duflo (2014) and Kizilcec et al. (2014) do not find positive impacts.
Relatedly, sending MOOC students a “planning prompt” improved course completion by 29 percent (Yeomans and Reich 2017).

Many educators firmly believe that discussion and interaction is a central component of education. But because MOOCs have thousands of students who generally access content at different times, regular discussions of the types that occur in classroom are rarely feasible. MOOC designers have attempted to at least partially address this problem by building discussion forums into MOOCs, but participation is often relatively low. Two experimental studies have evaluated efforts to increase participation in discussion forums. One study found insignificant or negative impacts from an email prompt (depending on the content of the email) (Kizilcec et al. 2014), while another found positive impacts on forum participation from asking participants to fill out a self-evaluation about forum participation (Lamb et al. 2015).

Another friction preventing efficient and equitable use of MOOCs may be “social identity threat,” the tendency of individuals—typically from marginalized social backgrounds—to “suffer from the cognitive burden of wrestling with feeling unwelcome while trying to learn and, therefore, underperform” (Kizilcec et al. 2017). Social identity threat has been shown to impair learning in a variety of ways. One recent set of RCTs evaluations tested the effects of writing exercises aimed at reducing social identity threat and found them to be effective in increasing persistence and completion among MOOC students from developing countries (Kizilcec et al. 2017). While this study focused on closing the gap between students from developed and developing countries, related interventions could also plausibly reduce social identity threat-driven gaps between advantaged and marginalized populations within the developed world.

5.2.3 Looking Forward

While these studies show some hope for behavioral nudges that encourage participation with online platforms and course completion of MOOCs, which may in turn help MOOCs better accomplish their goal of democratizing education, it should be noted that the literature is still nascent and there is a lot we still do not know. The online learning field is changing quickly, and new models that do not easily fit into the categories discussed here are springing up. For one, websites that offer more independent stand-alone modules—which allow for easier picking and choosing of content and use in supplementing other classes—are becoming increasingly important. The iconic website in this category is Khan Academy, which is currently undergoing several experimental evaluations, with studies either not yet completed or results not yet publicly available. Also popular in this space has been BrainPOP, which provides instructors with an expansive library of educational videos intended to be fun and engaging.

Another new development has been the rise of quasi-formal certification schemes, like Nanodegrees and MicroMasters, as alluded to above. These are certifications granted for completing sets of courses that are not formal degrees in the sense of college degrees, but that programs’ designers hope will increase their legitimacy and acceptance as real skill creators. Whether or not these quasi-formal certifications will be accepted as useful by employers and will come to take on some kind of labor market premium may become clear over the next few years. If employers had better ways of assessing skills during the hiring process, these programs could significantly expand education options. With regard to MOOCs, an important task for the research agenda will be to better articulate how exactly we judge success of MOOCs. For example, is access and course
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banerjee and Duflo</td>
<td><em>American Economic Review</em> (2014)</td>
<td>“Deadline effect” in the 14.73x: Challenges of Global Poverty MOOC—are students who register late less likely to do well or receive a certificate in the course?</td>
<td>Students who enrolled 1 day late were less likely to get a certificate and had lower grades.</td>
<td>Main reported effect sizes (not standardized): students who enrolled 1 day late were less likely to get a certificate (a reduction of 16.6 percentage points), and their grades were 10.7 percentage points lower.</td>
<td>Students registering within 15 days of deadline for 14.73x: Challenges of Global Poverty MOOC</td>
<td>42,314 students</td>
<td>Not randomized—regression discontinuity design</td>
</tr>
<tr>
<td>Banerjee and Duflo</td>
<td>Unpublished (2016)</td>
<td>1. Option to commit to structured study time 2. Self-efficacy messages 3. Tutoring services in groups of 20</td>
<td>Null effects on grades or engagement with courses</td>
<td>Null</td>
<td>Online course participants</td>
<td>19,694 students</td>
<td>Student</td>
</tr>
<tr>
<td>Davis et al. (2017)</td>
<td><em>Proceedings of the Seventh International Learning Analytics &amp; Knowledge Conference</em></td>
<td>A personalized feedback system that facilitates social comparison of current students with previously successful learners.</td>
<td>Positive effects on course completion rates</td>
<td>Main reported effect sizes (not standardized): 3.4 percentage point increase in course completion rates</td>
<td>Learners across 4 MOOCs provided by the Delft University of Technology on the edX platform</td>
<td>33,726 students</td>
<td>Students</td>
</tr>
</tbody>
</table>

(Continued)
TABLE 3B
MASSIVE OPEN ONLINE COURSES (MOOCs) (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davis et al. (2018)</td>
<td><em>Journal of Learning Analytics</em></td>
<td>MOOC-based Adaptive Retrieval Practice System, which delivers quiz questions from prior course units</td>
<td>Null effect on grades and obtaining course certificate</td>
<td>Null</td>
<td>Learners enrolled in edX geoscience course</td>
<td>1,047 students</td>
<td>Students</td>
</tr>
<tr>
<td>Kizilcec et al. (2014)</td>
<td><em>Open Education Europa eLearning Papers</em></td>
<td>“Collectivist,” “individualist,” or “neutral” emails sent to MOOC participants to encourage forum participation</td>
<td>1. Null effect on learners’ decision to contribute to the forum 2. Negative effect on contributions made by learners receiving the individualist encouragement and the collectivist message versus those receiving the neutral message</td>
<td>Main reported effect sizes (not standardized): 1. Null 2. Significantly lower contributions from those receiving individual encouragement and collectivist messages (exact magnitude not reported).</td>
<td>A subset of learners who enrolled in a MOOC on an undergraduate-level computer science topic at a major US university</td>
<td>3,907 students in first study, 7,522 students in second study</td>
<td>Student</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kizilcec et al. (2017)</td>
<td><em>Science</em></td>
<td>Mindset interventions addressing social identity threat using a “value relevance affirmation” exercise and a “social-belonging intervention”</td>
<td>1. Positive effects on persistence for learners in less-developed countries (LDCs) and null effects on persistence for learners in more-developed countries (MDCs); 2. In the second experiment, the social belonging intervention increased persistence for LDC learners without affecting persistence for MDC learners, and the affirmation experiment reduced persistence for MDC learners, but increased persistence for LDC learners</td>
<td>Main reported effect sizes (not standardized): 1. First experiment doubled persistence for learners in LDCs; null for learners in MDCs; 2. Second experiment: the social belonging intervention increases persistence for LDC learners without affecting persistence for MDC learners, and the affirmation experiment reduced persistence for MDC learners, but increased persistence for LDC learners (exact magnitude not reported)</td>
<td>2 samples: 1. Students from a computer science MOOC offered at Stanford 2. Students in a 6-week Harvard MOOC</td>
<td>2,286 students in first experiment, 1,165 students in second experiment</td>
<td>Student</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>-------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Lamb et al. (2015)</td>
<td>ACM (Association for Computing Machinery) Learning @Scale Conference Paper</td>
<td>Self-assessment questions aimed to improve forum participation for MOOC students: 1. A self-participation check, 2. Discussion priming and 3. Discussion preview emails</td>
<td>Positive effect on engagement and participation</td>
<td>Main reported effect sizes (not standardized): 4.2 more forum actions than users in the control group.</td>
<td>MOOC students in JusticeX, a HarvardX course</td>
<td>4,777 students</td>
<td>Student</td>
</tr>
<tr>
<td>Martinez (2014a)</td>
<td>EdPolicyWorks Working Paper</td>
<td>Emails informing students of their relative position in the course: 1. A “positive” one telling how many students recipients did better than, and 2. A “negative” one stating how other students outperformed the recipient</td>
<td>Positive effect on performance on subsequent quizzes from both “positive” and “negative” emails</td>
<td>Main reported effect sizes (not standardized): 2 percentage point increase from “positive” emails and 3 percentage point increase from “negative” emails</td>
<td>Students registered for a Coursera MOOC, Foundations of Business Strategy at the University of Virginia</td>
<td>7,924 students</td>
<td>Student</td>
</tr>
<tr>
<td>Martinez (2014b)</td>
<td>EdPolicyWorks Working Paper</td>
<td>E-mails on the negative correlation between procrastination and achievement</td>
<td>Positive effect on course completion.</td>
<td>Main reported effect sizes (not standardized): 16.85 percent increase in course completion.</td>
<td>Students from the third and fourth Foundations of Business Strategy MOOC at the University of Virginia</td>
<td>24,122 students from third MOOC, 5,675 from fourth MOOC</td>
<td>Student</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------------</td>
<td>-------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Patterson (2018)</td>
<td><em>Journal of Economic Behavior and Organization</em></td>
<td>1. A commitment device where students pre-commit to time limits on distracting Internet activities; 2. A reminder tool by time spent on distracting websites; 3. A focusing tool that allows students to block distracting sites whole on the course website</td>
<td>1. Commitment device: positive effect on time spent working on course, course grade, and course completion 2. Reminder and focusing treatments: null effects</td>
<td>1. 0.29 standard deviations 2. Null</td>
<td>MOOC participants in a Stanford OpenX course</td>
<td>657 students</td>
<td>Student</td>
</tr>
<tr>
<td>Yeomans and Reich (2017)</td>
<td><em>Proceedings of the Seventh International Learning Analytics &amp; Knowledge Conference</em></td>
<td>Open-ended planning prompts asking students to describe any specific plans they made to engage course content and complete assignments on time.</td>
<td>Positive effects on course completion</td>
<td>Main reported effect sizes (not standardized): 29 percent increase in course completion.</td>
<td>Students in 3 HarvardX MOOCs</td>
<td>1,792 students</td>
<td>Student</td>
</tr>
</tbody>
</table>
completion of all participants the best metric of success, given that many participants may not even aim to complete courses? More work is needed to hammer out what outcomes should be measured, beyond completion rates, to judge the success through closer investigations of where specifically they may add value to the education process. This will in turn require more nuanced study of students’ reasons for accessing MOOCs, and, more broadly, the role of MOOCs within the broader education field.

6. Technology-Enabled Behavioral Interventions

Next, we shift focus to education technologies that draw on the theory and practice of behavioral economics to guide students (and, in some cases, their family members and teachers) toward behaviors that are expected to facilitate greater academic achievement. The idea behind this approach is that people are subject to systematic biases in decision making and other psychological factors that lead to suboptimal outcomes (Kahneman 2011), like parents neglecting to read to children despite the best of intentions, or high school graduates not enrolling in college as a result of missing financial aid or registration deadlines. The behavioral insights literature was relatively slow to come to the education sector, but behavioral economics research in education has taken off over the past several years (Koch, Nafziger, and Nielsen 2015; Lavecchia, Liu, and Oreopoulos 2016; Levitt et al. 2016). Behavioral issues are especially important to think about in the context of education, since important long-run decisions are being made during a time when the brain’s ability to think of the future is not fully developed. While we all face challenges in making decisions involving uncertain long-run benefits and immediate costs, children and youth particularly struggle (Lavecchia, Liu, and Oreopoulos 2016).

With knowledge accumulating about how behavioral barriers get in the way of realizing better long-run outcomes, technology may increasingly be used to develop simple and inexpensive solutions to give individuals more support for making better choices. The predominant model of policy intervention that has arisen from the behavioral economics literature is the “nudge,” an “aspect of the choice architecture that alters people’s behaviors in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler and Sunstein 2008). Nudges can often be carried out inexpensively through technologies like text messages.

We identified 47 experimental papers studying technology-enabled behavioral interventions. These studies evaluated programs aimed at solving a wide variety of problems and drawing on a variety of techniques implemented at different points across the education life cycle, from giving parents ideas of how to practice reading skills in early childhood to reminding college students to submit for financial aid. In particular, we identified studies of interventions across four clusters: seven on encouraging parental engagement in learning activities, 13 on attempting to improve information flows in postprimary and secondary school, 19 on encouraging success in transitioning to and through college, and 15 on psychosocial interventions. Information on the studies is presented in tables 4A–4D. The studies show strong promise in each of these areas.

While technology access interventions aim to create an enabling environment for learning, and CAL and online courses attempt to directly improve the efficiency of academic and cognitive skills investments, behavioral interventions typically seek to overcome hurdles in the education process brought on by gaps in certain types of noncognitive skills. The term “noncognitive” skills refers to a wide range of social and emotional skills,
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortes et al. (2018)</td>
<td>NBER Working Paper</td>
<td>Text messaging program to nudge parents of kindergarteners to engage in literacy activities with children (continuation of Doss et al. 2018). Varies content and number of text messages per week.</td>
<td>Negative effect from having too few messages (1/week) and too many messages (5/week) compared to baseline (3/week)</td>
<td>1. 0.25 standard deviation decrease from 5/week texts compared to 3/week 2. 0.19 standard deviation decrease from 1/week text.</td>
</tr>
<tr>
<td>Doss et al. (2018)</td>
<td>NBER Working Paper</td>
<td>Text messaging program to nudge parents of kindergarteners to engage in literacy activities with children (continuation of York and Loeb 2014). Includes additional differentiated/personalized treatment arm</td>
<td>1. Positive effects on reading level 2. Positive effect on parental engagement in literacy activities by compared to the control group 3. Null effects for other treatment arm</td>
<td>1. 50 percent more likely to read at a higher level compared to the general group (effect size not standardized) 2. 0.31 standard deviation increase in literacy activities 3. Null</td>
</tr>
<tr>
<td>Kraft and Monti-Nussbaum (2017)</td>
<td>ANNALS of the American Academy of Political and Social Science</td>
<td>Parents texted to encourage to engage in activities to counteract summer learning loss</td>
<td>Positive effect on reading comprehension, more than compensating for summer learning loss</td>
<td>0.21 to 0.29 standard deviations increase for third and fourth graders on reading comprehension</td>
</tr>
</tbody>
</table>

Sample: Preschool students and families in Dallas, Texas
Sample: Kindergarten students and families in California
Sample: 183 US families (elementary school kids)
Sample: 232 students in 183 families

Unit of randomization: Parent

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mayer et al. (2019)</td>
<td>Journal of Human Resources</td>
<td>Texting program to promote learning engagement of Head Start parents</td>
<td>Positive effect on reading application usage</td>
<td>1. 1.01 standard deviation increase in minutes read</td>
<td>Parents of children in subsidized preschool programs in Chicago</td>
<td>169 parents</td>
<td>Parents</td>
</tr>
<tr>
<td>Meuwissen et al. (2017)</td>
<td>Minnesota Children's Museum and the Center for Early Education and Development, University of Minnesota</td>
<td>Text2Learn, a mobile phone texting program for low-income parents of preschoolers.</td>
<td>1. Positive effect on parental engagement in more literacy activities with their children 2. Null effects on community resources usage, such as libraries, or changes in attitudes about literacy</td>
<td>Main reported effect sizes (not standardized): 1. Parents reported engaging in more literacy activities with their children after receiving the texts (exact magnitude not reported) 2. Null</td>
<td>Parents of preschool children in Minnesota</td>
<td>110 parents</td>
<td>Parents</td>
</tr>
<tr>
<td>York and Loeb (2019)</td>
<td>Journal of Human Resources</td>
<td>Text messaging program to nudge preschool parents to engage in literacy activities with children</td>
<td>Positive effect on engagement in literacy activities, parental involvement at school, and learning gains</td>
<td>1. 0.22–0.34 standard deviation increase in literacy activities 2. 0.13–0.19 standard deviation increase in parental involvement at school, and 3. 0.21–0.34 standard deviation increase in learning gains</td>
<td>Parents of 4-year-olds in California</td>
<td>440 parents</td>
<td>Parent</td>
</tr>
</tbody>
</table>
most of which are well outside the scope of this review. We refer instead to the much narrower set of noncognitive skills—such as time management, motivation, and resilience—that can conceivably be affected by nudges, like reminders, or messages containing small bits of information and encouragement. While these nudge interventions may or may not actually directly help to develop the noncognitive skills in question, they at least attempt to compensate for gaps to lower their potential for inhibiting the education process. In the remainder of this section, we review the rationale and evidence underlying each of the four clusters enumerated above in turn.

6.1 Encouraging Parental Learning Engagement during Early Childhood

6.1.1 Background, Context and Mechanisms of Impact

Research suggests that an effective strategy to improve educational outcomes is for parents to engage in learning activities with their children (Price 2010, Sénéchal and Lefèvre 2002). But parents report spending less time on these activities than might be expected in light of the possible benefits. The problem of low engagement is particularly acute among disadvantaged households, a pattern that may reinforce broader disparities in educational outcomes (Guryan, Hurst, and Kearney 2008; Kalil 2015; Lee and Bowen 2006). Policy makers have found cost-effective responses elusive, with even expensive and resource-intensive programs turning up modest results (York, Loeb, and Doss 2019). Yet because young children spend a great deal of time at home, school-based programs cannot substantially substitute for engagement “unless they are very intensive, extensive and expensive” (Mayer et al. 2015). This dilemma has inspired a growing literature that explores whether and how behavioral interventions might contribute toward reducing disparities in family engagement.

Why might nudges be expected to increase parental learning engagement within disadvantaged households? After all, behavioral interventions are unlikely to substantially address resource constraints like the time scarcity faced by low-income parents. However, the behavioral economics literature suggests that cognitive constraints as well as resource limitations lead to underinvestment. Even when cognitive burdens themselves are aggravated by resource constraints, small adjustments in the decision structures that people face can help to correct these biases and move them toward more optimal behavior (Thaler and Sunstein 2008).

So, in the present context, a behavioral economics perspective would indicate potential benefits from reminders and instructions inspiring and guiding parents toward more productive engagement. Behavioral interventions are likely to be helpful to the extent that the current lack of parental involvement in learning activities reflects frictions like limited mental bandwidth (Schilbach, Schofield, and Mullainathan 2016) and procrastination (Thaler and Benartzi 2004) rather than the reflecting well-reasoned cost–benefit decisions or insurmountable resource barriers. In the context of our framework, interventions within this category could help to compensate for deficits in noncognitive skills associated with procrastination.

6.1.2 Experimental Evidence

We identified six experimental evaluations of technology-based interventions aiming to increase the quantity and quality of time spent by parents practicing skills with their preschoolers (Cortes et al. 2018, Hurwitz et al. 2015, Mayer et al. 2019, Meuwissen et al. 2017, York and Loeb 2014), kindergarteners (Doss et al. 2018), or first–fourth graders (Kraft and Monti-Nussbaum 2017). All
of the programs studied relied centrally on sending text message reminders to parents, and all found positive results.

Ready4K—a preschool literacy program implemented in San Francisco—was the earliest experimentally evaluated, technology-based intervention to leverage this rationale to improve parental learning engagement. The program sent parents three text messages per week with tips and encouragement to engage in literacy activities (York and Loeb 2014). The behavioral logic that guides Ready4K suggests that “the complexity of parenting may overwhelm some parents, leading them to underinvest in their children” (York and Loeb 2014). Furthermore, literacy activities constitute a case of “delayed gratification,” necessitate “interrupting the status quo,” and are often overcome by “limited attention” (York and Loeb 2014). So the program sends suggestions of small, easy tasks that parents can carry out without feeling overwhelmed; provides encouragement to sustain parents’ investment in longer term gratification; provides tips for integrating the activities into daily life so that the status quo barrier can be overcome; and addresses attention constraints by regularly reminding parents.

The study found an impact of 0.29 standard deviations of the program on a composite score for “global early literacy parenting” measuring activities like reading to a child, pointing out words that rhyme, and taking the child to a library or museum (York and Loeb 2014). The study also found effect sizes ranging from 0.21 to 0.34 standard deviations on the Phonological Awareness Literacy Screening (York and Loeb 2014). The fact that the program led to an increase in specific literacy tasks but not general ones suggests that the impact was likely generated by the program’s provision of particular, manageable tasks, rather than reminding parents to engage in activities they might have engaged in anyway. The effect sizes detected are impressive given the exceedingly low costs of the intervention, at less than a dollar per family (York and Loeb 2014). To build on the promising evidence, Ready4K is now exploring whether the program can improve math performance and executive functioning.

Ensuing research has attempted to work toward untangling the specific mechanisms underpinning the effectiveness of this type of intervention, as well as better understanding potentially differing effects across subgroups. The only remaining intervention of this type for preschoolers to have been experimentally evaluated took place within Midwestern Head Start and Early Head Start centers. The program provided all sample households with tablets containing numerous children’s books (Mayer et al. 2019). The treatment group additionally received three nudges—daily text message reminders to read to the kids, a goal-setting tool that asked the parents to set reading goals and reported back on whether these goals were met, and social rewards, specifically congratulatory texts or cartoons when goals were reached (Mayer et al. 2019). Following the six-week study period, the group receiving the behavioral interventions used the tablet a full standard deviation more than parents who did not. They read more than twice as many books to their children, with control group families reading an average 14.8 books during the six-week intervention period while treatment group families read an average of 31.4 books. The fact that effects were stronger for more present-oriented than for more future-oriented parents suggests that the mechanism was behavioral—in this case, helping to correct for temporal bias—rather than simply arising from the introduction of new information (Mayer et al. 2019).

As children progress from preschool to kindergarten and then first grade, they tend to spend larger shares of their time at school. To what extent might programs like the ones described above prove effective beyond
preschool? Two interventions were recently experimentally evaluated that adopt a similar model, but for kindergarteners (Doss et al. 2018) and first–fourth graders (Kraft and Monti-Nussbaum 2017) instead of preschoolers. The kindergarten intervention was an extension of Ready4K with the evaluation including the same preschool sample as the children entered kindergarten, along with additional San Francisco kindergarteners (Doss et al. 2018). Beyond trying to replicate the same intervention within a kindergarten context, a second treatment arm was also added that sent parents “personalized” and “differentiated” texts. Texts to parents in this second treatment arm contained child-specific information and sent recommendations for tasks matching the child’s level. Interestingly, the researchers found the original treatment that had been effective in preschool showed no significant effects in kindergarten. However, the personalized and differentiated text messages did show substantial benefits, with children whose parents received the treatment “50 percent more likely to read at a higher level” (Doss et al. 2018).

Finally, the most recent intervention, falling into this category, to undergo experimental evaluation extended the idea of texting parents to encourage engagement in literacy activities to the first–fourth grades. Recognizing that elementary students spend more time engaged in school throughout the year, this intervention targeted a specific point of friction within the elementary education process—“summer reading loss”—the tendency of elementary students to fall behind in their reading skills because of the gap in practice they experience during the summer. Parents in the treatment group were sent 18 text messages across July and August with information on the importance of summer reading, as well as affordable and accessible resources to practice reading and tips on parental engagement. The study finds that the texting intervention improved reading comprehension intervention scores for students in the treatment group by 0.21–0.29 standard deviations (Kraft and Monti-Nussbaum 2017).

6.1.3 Looking Forward

Given the enormous importance of early childhood skill development for the entire educational life course, coupled with the central role of parents and caretakers in this skill development process, behavioral nudges to encourage educational interactions with children represent a vital area for continued experimentation and scale-up. New work in this area should explore several directions. One important question to address is when and how customization should be factored in. Of the studies discussed here, only one study (Doss et al. 2018) involves a customized treatment arm, and this arm shows strong effects. While non-customized interventions show positive effects at the preschool (Cortes et al. 2018; Mayer et al. 2019; York, Loeb, and Doss 2019) and third and fourth grade levels (Kraft and Monti-Nussbaum 2017), it is noteworthy—although of course far from conclusive—that non-customized interventions show no effects for kindergarteners (Doss et al. 2018) and first–second graders (Kraft and Monti-Nussbaum 2017). Customization can add costs, but may lead to greater cost effectiveness for particular ages and particular subjects or learning tasks—sorting this out is an empirical question that will require multi-armed experiments. Moreover, with the growing sophistication of artificial intelligence, there is potential for customized text messaging platforms to be developed at a lower cost.

Second—as suggested by the large effect sizes found in the Head Start study (Mayer et al. 2019)—a potentially promising route is to overlay messaging interventions onto other interventions providing resources. Program designers have often struggled with the challenge of distributing potentially
valuable resources, only to come up against low usage rates. Similarly, it is easy to imagine that, while tips and reminders might be helpful, parents would be more likely to heavily engage with a robust set of resources at their disposal, like books to read to their children. It is thus worth exploring how messaging interventions could complement other resource distribution programs.

Finally, it will be important for researchers to explore the effects of delivery mechanisms other than text messages. While text messages are still relatively novel for many parents and are likely more noticeable than email, this may gradually change as organizations increasingly leverage text messages for promotions and parents generally become more inundated with text messages. Messages can also be sent through WhatsApp, social media, and other platforms. Experiments could compare the effects of different platforms—or combinations of platforms—in multiple treatment arms.

6.2 Improving School–Parent Information Flows

6.2.1 Background, Context, and Mechanisms of Impact

As children get older, the role of parents shifts away from practicing skills with their kids directly and towards encouraging the kids to put more effort into school. So, behavioral interventions for middle and high schoolers tend to focus on sending parents information on their kids’ performance—for example, updates on grades, attendance, and behavior—to nudge the parents toward providing this encouragement. If parents are constrained by a gap in information on how hard their children are working or how well they are performing, and if children are not already expending maximum effort, then closing these gaps may provide parents with the opportunity to apply that alchemical combination of guidance, pressure, and support that constitutes parenting. This issue may be especially important for low-performing schools, which already exhibit lower rates of communication satisfaction from parents and where parents may be relatively more constrained in their ability to absorb monitoring costs (Bergman 2015).

Within our framework, improving parent–child information flows can help to build or compensate for noncognitive skill deficits. From the perspective of parents, many of the same dynamics mentioned in the preceding section may hold here. Parents may be aware of the importance of obtaining information on their children’s performance and providing guidance accordingly, but they may procrastinate given the many important topics competing for attention, or they may feel overwhelmed and have trouble getting started. Children also face many competitors for their attention and are notoriously wont to prioritize social identity over academic performance (Akerlof and Kranton 2002), leading them to procrastinate or forego schoolwork altogether if not sufficiently prompted by parents. To the extent that a child’s academic standing is clear and salient to parents, this could spark the action needed to cultivate the child’s conscientiousness, which is one of the traits that has been identified as closely associated with academic and professional success (Almlund et al. 2011, Almund 2011).

6.2.2 Experimental Evidence

We identified 13 RCT-based studies evaluating technology that seeks to improve the flow of information in postprimary and secondary school, the majority of which focus on improving school–parent information flows (Balu, Porter, and Gunton 2016; Bergman 2015, 2016; Bergman, Denning, and Manoli 2019; Bergman and Chan 2017; Bergman and Rogers 2016; Fryer 2016; Kraft and Dougherty 2013; Kraft and Rogers 2015; Kraft and Monti-Nussbaum 2017; Rogers...
These programs followed two main approaches: first, sending information to parents that was generated as part of regular school records (like grades and attendance) and, second, having teachers send personalized messages to parents. Overall, these studies have found positive results, indicating a potentially fruitful set of opportunities.

The majority of the school–parent information flow interventions that have been experimentally evaluated fall into the first of the two categories listed above. The first intervention in this category to be experimentally evaluated was a program aimed at middle and high school students at a single public school in a low-income neighborhood of Los Angeles (Bergman 2015). Parents whose children were in the treatment groups were notified when their children missed attending class or missed an assignment through text messages, phone calls, and e-mails. Following the semester-long intervention, students in the treatment group had earned GPAs and standardized math test scores that were about 0.20 standard deviations over those of the control group (Bergman 2015). An evaluation of a similar intervention—Papás al Día (“Parents up to Date”), carried out in two low-income municipalities of Santiago, Chile—also finds positive results, including a 0.09 standard deviation improvement in math grades, a reduction in bad behavior, and positive spillover effects within classes (Berlinski et al. 2016).

While these two interventions sought to channel existing information on students’ performance to parents rather than generating new information, both were somewhat labor intensive, requiring substantial manual data entry. More recent interventions have tended to automate the process to the greatest extent possible to cut down on costs. One recent experimental study evaluated the effects of an automated school–parent information program on a sample of 22 middle and high schools in a district of West Virginia (Bergman and Chan 2017). This program automatically pulled information from the schools’ student information system and texted it directly to parents. Parents received weekly texts stating the number of classes and/or assignments that students had missed, as well as monthly texts if their child was averaging below 70 percent on any class (Bergman and Chan 2017). Because of the automation, the intervention was extremely cheap, with 32,000 text messages totaling to only $63 and training coming down to $7 per student (Bergman and Chan 2017). The study showed impacts that were very impressive, given the low costs of the intervention: the treatment group saw a 39 percent reduction in failed courses and an 18 percent increase in class attendance, meaning that the treatment group attended 50 more classes on average and had a 0.10 standard deviation improvement in GPA (Bergman and Chan 2017). Interestingly, the data suggest that parents already had a good idea of their children’s final grades, but the program reduced parents’ underestimation of the number of assignments their kids were missing, which likely helped to better target the pressure they placed on their kids to increase effort (Bergman and Chan 2017). The strongest benefits went to those with below-average GPAs, who saw a reduction in class failures of 0.9 classes, an increase in attendance of 64 classes, and a GPA increase of 0.26 points out of a 4.0 scale (Bergman and Chan 2017).

In contrast, another fully automated intervention that focused exclusively on attendance showed no evidence of improving attendance rates (Balu, Porter, and Gunton 2016). Here, parents of New York City Public School students received automated text messages on each day their student did not show up for school, in addition to weekly attendance reports. This lack of impact may have arisen from the intervention’s exclusive...
focus on attendance, its location in New York City (which may be more saturated with automated information flows than most other environments), or something more contingent and specific to the intervention in question. However, it is noteworthy that this study—the only one to focus on an attendance-only intervention—was also the only one to show no effects. Further research will be needed to explore the possibility that it is effort within school, rather than attendance, that constitutes a binding constraint on parent–child information flows in most high school settings.

Two recent studies have highlighted an important qualification to the line of research just described (Bergman 2016). While technologies that improve school–parent information flows may be effective in improving education, these effects will be heavily mediated by the extent to which the technologies are actually used. For instance, one recent study showed a letter and phone call prompting students to access an online system containing attendance and grades significantly increased rates of access and ultimately resulted in a GPA increase of 0.10 points (Bergman 2016). Another program—this one conducted in a dozen Washington, DC middle and high schools—offered text message updates of the kind mentioned above, but varied in how the program was implemented (Bergman and Rogers 2016). Three treatment groups—one that received a text instructing them on how to sign up online for the service, one that received a text inviting sign-up through a text message response, and one that automatically enrolled parents in the texting program but gave them the opportunity to opt out—were contrasted with a control group that did not receive any prompt to sign up for the texting service. Only 1 percent of participants in the first group and 8 percent in the second group signed up, while only 4 percent in the automatic enrollment group chose to opt out. This massive difference in adoption shaped the effectiveness of the texting program in generating academic performance outcomes: while no significant effects on performance outcomes emerged from the first two treatment groups, the automatic enrollment group saw improvements in GPA by roughly a quarter to a third of a letter grade, and reduced class failure by roughly a fifth to a quarter (Bergman and Rogers 2016). These lessons on the importance of encouragement and especially opt-in systems to promote technology adoption are relevant to a broad range of ed-tech applications, but are mentioned here since they were evaluated in reference to school–parent communication intervention.

The interventions discussed so far in this subsection attempt to transfer already-existing information to parents. Another approach that has been experimentally evaluated in the context of two separate interventions has teachers communicate personalized messages to parents. The first experimentally evaluated intervention falling into this category took place during a required summer program in a Boston charter school (Kraft and Dougherty 2013). Parents received two communications per day for five consecutive school days—a phone call from an English teacher and a text message from a math teacher. The intervention improved engagement as measured by three variables: homework completion, participation, and number of instances in which teachers had to direct students’ attention back to the topic at hand (Kraft and Dougherty 2013). Qualitative evidence suggests that this effect occurred through three mechanisms: improving relationships between students and teachers, expanding parental involvement, and increasing students’ motivation (Kraft and Dougherty 2013).

The second intervention in this category took place “during a traditional summer
school program offered by a large urban school district in the Northeastern United States" (Kraft and Dougherty 2013). Here, teachers themselves wrote out one-sentence messages that were then sent to parents weekly by research assistants through text message, phone, or email (Kraft and Dougherty 2013). Two separate program variations were given: one consisting of “positive” messages about what the student was already doing well, and the other consisting of “improvement” messages about areas that the student could use work on. Averaging across the two treatment arms, inclusion in the program led to an increase in the success rate of students passing the class and obtaining the credit, up 6.5 percentage points from an 84.2 percent passing rate in the control group. Interestingly, the impact estimate is substantially higher for the improvement treatment arm, although the experiment lacks the power to detect significance in this difference (Kraft and Dougherty 2013). The program seems to work not by increasing the amount of time parents spend talking with their kids about school, but rather by directing the content of these conversations. The program also seems to have led to the unintended consequences of lower student perceptions of their own performance, and weaker student–teacher relationships as reported by teachers also (Kraft and Dougherty 2013). Perhaps the best of both variations could be captured by sending messages that include actionable steps as in the “improvement” version, but are more positive in tone.

6.2.3 Looking Forward

Overall, other than the lack of impact generated by the New York City attendance program, interventions that seek to improve school–parent information flows seem highly promising. They may be especially cost effective when automated and drawing on existing school administrative data. Opportunities within this latter area may increase as schools increasingly turn to computerized grading and feedback systems.

It would be productive to extend research on parent–child information flows in several directions. The interventions studied so far have been focused primarily on providing parents with information about their children’s performance, but without any guidance on what to do with this information. Messages could be coupled with actionable tips about how to encourage struggling students, or resources like informative websites or contact information for school counselors. And they can be imbued with tiers, or importance rankings, in order to catch a parent’s attention if something especially worrying is going on (e.g., if the student is on the verge of failing a class). As in the preceding section, experimentation with more mechanisms of communication beyond text messages would be worthwhile. And sending children information or alerts relating to their performance—with the understanding that their parents will receive a message if things do not change—could also increase student effort without even necessarily involving the parents.

6.3 Transitioning to and Succeeding in College

6.3.1 Background, Context, and Mechanisms of Impact

Another area of focus for technology-based nudge interventions in the education sector has been the challenge of transitioning to and making it through college. The behavioral economics literature suggests that people—and especially children, adolescents, and young adults—tend to rely heavily on routines, and the transition to college requires students to break from routine (Laveccchia, Liu, and Oreopoulos 2016). The behavioral literature has also documented the paralyzing effect of too much information and too
<table>
<thead>
<tr>
<th>Author, Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balu, Porter, and Gunton (2016) MDRC</td>
<td>Automated text messages to parents of high school students informing about absence</td>
<td>Null effect</td>
<td>Null</td>
<td>New York City high school students</td>
<td>3,957 students</td>
<td>Student</td>
</tr>
<tr>
<td>Bergman (2015) CESifo Working Paper</td>
<td>Automated texts to parents about performance</td>
<td>1. Positive effects on high school GPA 2. Decrease in missing final exam project 3. Positive effect on math standardized exam scores 4. Null effects on English</td>
<td>1. 0.19 standard deviation increase in high school GPA 2. 7.5 percentage point decrease of missing final exam project (not standardized) 3. 0.21 standard deviation increase for math standardized exam scores 4. Null for English</td>
<td>Students in grades 6–11 in Los Angeles</td>
<td>462 students</td>
<td>Student</td>
</tr>
<tr>
<td>Bergman (2016) CESifo Working Paper</td>
<td>Learning Management System (parents have access to an online portal with child’s classes, grades, assignments, etc.)</td>
<td>1. A quarter of parents use online portal 2. Adoption follows an S-shape 3. Significant spill- overs occur along intensive but not extensive margins 4. There is evidence student grades improve as a result</td>
<td>Main reported effect sizes (not standardized): 24 percent of parents log into the system</td>
<td>15 US school districts operating learning management company; 2-stage experiment providing families their account information in schools across 3 districts.</td>
<td>59 schools</td>
<td>School</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>-------------</td>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>Bergman and Chan (2017)</td>
<td>CESifo Working Paper</td>
<td>Automated texts to parents about performance</td>
<td>1. Positive effect on reducing course failures, GPA, and attendance 2. Null effects on state math and reading scores 3. Positive effect on in-class exam scores</td>
<td>1. 40 percent reduction in course failures (not standardized) 2. 0.10 of a point GPA increase for middle school students and .25 of a point GPA increase for high school students (not standardized) 3. 17 percent increase in class attendance (not standardized) 4. Null in state math and reading scores 5. 0.10 standard deviation increase on in-class exam scores</td>
<td>22 middle and high schools in Kanawha County Schools in West Virginia</td>
<td>1,137 students</td>
</tr>
<tr>
<td>Bergman, Edmond-Verley, and Notario-Risk (2018)</td>
<td>Economics of Education Review</td>
<td>Providing regular information to families about their child’s academic progress in one arm and supplementing with home visits on skills-based information in a separate arm</td>
<td>1. Positive effect on GPA and math and reading test scores for the treatment arm with home visits 2. Positive effects on retention for both groups during the year</td>
<td>1. Information only: 0.13 standard deviations in GPA; null on math and reading test scores 2. Information + skills: 0.8 standard deviations increase in GPA; 0.13 and 0.12 standard deviation increase in math and reading test scores 3. 4 percentage points less likely to leave the district (pooled)</td>
<td>Students’ families from 3 participating schools in an urban, Midwestern school district</td>
<td>1,121 students’ families</td>
</tr>
</tbody>
</table>
TABLE 4B
TECHNOLOGY-ENABLED BEHAVIORAL INTERVENTIONS (PRIMARY AND SECONDARY) *(Continued)*

<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergman and Hill (2018)</td>
<td><em>Economics of Education Review</em></td>
<td>Publishing teacher ratings online</td>
<td>1. High-performing students sort into classrooms with highly rated teachers</td>
<td>A 1 standard deviation increase in a teacher’s value added causes the math test scores of incoming students to be three tenths of a standard deviation higher on average than those of a similar teacher whose rating is not published</td>
<td>Third–fifth grade teachers in Los Angeles</td>
<td>3,089 teachers</td>
<td>Not randomized—regression discontinuity design</td>
</tr>
<tr>
<td>Bergman and Rogers (2016)</td>
<td>Society for Research on Educational Effectiveness Spring 2016 Conference Paper</td>
<td>Text message to parents regarding their child’s academic performance, including grades, upcoming tests, and missing assignments</td>
<td>1. Positive effect on grades for those assigned to opt out group 2. Overall, positive effect on grades</td>
<td>1. 0.06 standard deviation increase from being assigned to opt out for term 3 and 0.04 standard deviation increase for term 4 2. 0.05 standard deviation increase in grades overall for terms 3 and 4</td>
<td>Middle and high school students in 12 US schools</td>
<td>6,976 students</td>
<td>Student</td>
</tr>
</tbody>
</table>

*(Continued)*
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bursztyn and Jensen (2015)</td>
<td>Quarterly Journal of Economics</td>
<td>Two interventions: 1. Performance leaderboard into computer-based high school courses 2. Complimentary access to an online SAT preparatory course; sign-up forms differed randomly across students only in whether they said the decision would be kept private from classmates</td>
<td>1. Negative effect on student performance; the decline appears to be driven by a desire to avoid the leaderboard 2. In non-honors classes, sign-up was 11 percentage points lower when decisions were public rather than private; honors class sign-up was unaffected</td>
<td>Main reported effect sizes (not standardized): 1. 24 percent performance decline 2. 11 percentage point decrease in sign-ups in non-honors classes when decisions were public rather than private; null in honors class sign-up.</td>
<td>Study 1: students across more than 100 schools in Los Angeles Study 2: 26 classrooms across 4 schools in Los Angeles</td>
<td>5,000 students in first study, 825 students in second study</td>
<td>Student</td>
</tr>
<tr>
<td>Fryer (2016)</td>
<td>Journal of Public Economics</td>
<td>Provided free cellular phones and daily information about the link between human capital and future outcomes via text message in one treatment and minutes to talk and text as an incentive in a second treatment</td>
<td>1. Positive impact on students' reported beliefs about the relationship between education and outcomes 2. Null effects on student effort, attendance, suspensions, or state test scores; evidence of positive impact on college entrance exams 4 years later</td>
<td>Main reported effect sizes (not standardized): 1. 23 percent more questions answered correctly on relationship between human capital and life outcomes 2. Null</td>
<td>Students in sixth and seventh grades in Oklahoma</td>
<td>1,907 students</td>
<td>Student</td>
</tr>
</tbody>
</table>


### TABLE 4B

<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
</table>
2. Decreased instances in which teachers had to redirect students’ attention to the task at hand  
3. Positive effect on class participation rates                                                                                                           | Main reported effect sizes (not standardized):  
1. 40 percent increase in homework completion  
2. 25 percent decrease in instances in which teachers had to redirect students’ attention  
3. 15 percent increase in class participation rates                                                                                                        | Rising sixth and ninth grade students in Boston, Massachusetts                                                   | 140 students                           | Student |
<p>| Kraft and Rogers (2015)  | <em>Economics of Education Review</em>                  | Parents texted on student behavior/performance                               | Decreased the percentage of students who failed to earn course credit                                                                                                                                                  | Main reported effect sizes (not standardized): 41 percent reduction in failure to earn course credit | High school students in Northeastern United States                                                                   | 435 students                          | Students |</p>
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>McGuigan, McNally, and Wyness</td>
<td>Journal of Human Capital</td>
<td>Information campaign about the costs and benefits of pursuing post-compulsory education</td>
<td>Positive effect on understanding of university education financing and secondary school retention; null impact on plans to apply to and enroll in college</td>
<td>Main reported effect sizes (not standardized): 1. 41 percent increase in knowing when university fees are paid 2. 25 percent increase in knowing the student loans are a cheaper/better way to borrow 3. 16 percent increase in staying in full-time education after age 16 4. Null impacts on the perceived importance of family constraints on college enrollment 5. Null impacts on plans to apply to and enroll in college</td>
<td>Year 10 students in England</td>
<td>6,614 students in 54 schools</td>
<td>School</td>
</tr>
<tr>
<td>Rogers and Feller</td>
<td>Nature Human Behavior</td>
<td>1 of 3 personalized message information treatments throughout the school year</td>
<td>Positive effect on reduction of chronic absenteeism comparably across all grade levels</td>
<td>Main reported effect sizes (not standardized): 10 percent reduction in chronic absenteeism</td>
<td>Low-income households across 203 US schools</td>
<td>28,080 households</td>
<td>Households</td>
</tr>
</tbody>
</table>
many choices, and the transition to college is fraught with these as well (Lavecchia, Liu, and Oreopoulos 2016). In the context of our framework, behavioral interventions can help to overcome or compensate for noncognitive skill deficits that could otherwise hamper a student’s entire academic career. For instance, procrastinating—or, for that matter submitting to information overload paralysis—can harm a student academically if it causes him or her to turn in a late or low-quality assignment. But missing an important admissions or financial aid deadline can mean consequences that are orders of magnitude more significant, for example delaying college, ending up in a low-quality institution, or even foregoing college altogether. The interventions discussed in this subsection aim to provide students with tools of varying intensity to avoid such problems. In particular, we review evidence on four main types of college-related behavioral interventions: information campaigns, nudges to complete important tasks, intensive application assistance, and college advising.

6.3.2 Experimental Evidence

Our paper reviews 19 studies (included in table 4C) that evaluate the use of technology-enabled behavioral interventions to support the transition to and success in college. First, several interventions have sought to leverage information technology to inexpensively provide students with more college-related information. On one hand, two relatively minimalistic interventions in the United States generated no impact. One of these—tested in a field experiment with a sample of over a million prospective and enrolled college students in Texas—sent one e-mail and one letter containing information about higher education tax credits, but those who received these showed no more likelihood of applying to or enrolling in college than those who did not (Bergman, Denning, and Manoli 2019). Another intervention, conducted in a single public university, emailed letters to students explaining their current financial aid package and associated plans, but this information too had negligible effects (Darolia and Harper 2018).

On the other hand, two information interventions implemented in Canada and Chile, respectively, found positive effects. The first of these interventions showed videos on the benefits of higher education to students in disadvantaged Toronto high schools and allowed the students an opportunity to try out a financial aid calculator. Students who participated in the program reported more favorable views of higher education (Oreopoulos and Dunn 2013). The other program sent eighth graders in metropolitan Santiago, Chile, DVDs containing practical information on higher education financing. Participants not only showed greater knowledge of financial aid, but also were more likely to enroll in college preparatory high schools, and also exhibited postsecondary attendance rates that were 8.8 percent higher (Dinkelman and Martínez 2014). This latter intervention is also unique among programs that have been experimentally evaluated in that it targets higher education at the eighth-grade level, which could allow more time for participants to plan for college.

More research would be needed to definitively explain why the first two interventions showed no results and the next two did. However, one possibility is that the first two relied on letters, which the intended beneficiaries may or may not have found compelling or even read. Perhaps video—the primary medium used in the second two interventions—was more effective at capturing students’ attention. For the Texas tax incentive experiment, a simpler interpretation—and the one offered by the authors—is that the tax incentives themselves did not figure into students’ decision making, and so more information on the incentives did not help.
Another approach to supporting the transition to college has been through nudge campaigns. Although the term “nudge” is applicable to interventions discussed throughout this section, here we use the term “nudge campaigns” to refer to interventions providing sustained efforts to guide, encourage, and/or remind program participants about one or more aspects of college success. Five recent studies all suggest that nudge campaigns can be effective in improving decisions and task fulfillment surrounding financial aid and college matriculation and enrollment.

Of these, three interventions attempted to encourage better-informed financial aid decisions. One program sent students at a large community college in Baltimore County eight text messages over a period of several weeks prompting them to make more “active” financial aid decisions. The intervention resulted in a 3.1 percentage point reduction among students who received the text messages in accepting unsubsidized Stafford loans, and those who still did accept the loans borrowed less. Results were strongest among students showing less financial literacy and with more debt. The study also produced some evidence that the text messages led students who had attained marginal academic success to leave school earlier (Barr, Bird, and Castleman 2016). Another program sent text messages to college freshmen who, as high school students, had worked with a Massachusetts-based education nonprofit called uAspire. The messages encouraged students to refile the Free Application for Federal Student Aid (FAFSA) for their sophomore year and found an increase of nearly 14 percentage points on continuous enrollment through sophomore year among students attending community colleges (those attending four-year universities already had high rates of continuous enrollment) (Castleman and Page 2016).

Most recently, the largest experimentally evaluated FAFSA nudge campaign to date sent three versions of a message to low-income and first-generation students filling out the Common Application encouraging them to apply early for the FAFSA. One version provided specific planning structure, one gave information on the human capital returns to college, and one attempted to advocate productive identities. No effects were found for the latter two frames, but the planning message led to a 1.1 percentage point increase in college enrollment among all recipients and 1.7 percentage points for first-generation college students (Bird et al. 2017). In addition to supporting task completion related to financial aid, one nudge campaign has been experimentally shown to reduce “summer melt,” the phenomenon whereby students who are admitted to and indicate a decision to attend a particular college do not actually complete the matriculation process or do not actually show up for classes (Castleman and Page 2015).

A nudge campaign may be sufficient to induce students to think through financial aid decisions and remind them to do the right paperwork on time to enroll in and get through school. However, it is perhaps less likely that nudges would be effective at getting a student to fill out an admissions or financial aid application in the first place—this is a much more daunting task. We identified evaluations of two programs that leveraged technology for more intensive application assistance and support (Bettinger et al. 2012, Oreopoulos and Ford 2019). In the first instance of these programs, families with a college-age child who were filing their taxes at H&R Block were given the opportunity to quickly file their FAFSA at the same time. This was possible as a result of a software program designed to automatically feed data from the tax entry system into the FAFSA, collecting additional FAFSA questions not covered during the course of
the regular tax filing in ten or so minutes following the filing. College enrollment of high school seniors with parents receiving the treatment increased by 8 percentage points (Bettinger et al. 2012). The program LifeAfterHighSchool, on the other hand, focused on providing support for the admissions process directly to students by incorporating relevant activities into the high school curriculum (Oreopoulos and Ford 2019). The program aimed to ensure that every senior in participating high schools graduate with a college program offer of acceptance and a financial aid package. The program consisted of workshops involving interactive activities, for instance having students enter their grades into a computer program, which would then generate a list of local programs in their area for which they would likely be accepted if they applied. In addition to large gains in application rates, college enrollment increased by about 9 percentage points among the seniors who had not been taking any university-track courses (Oreopoulos and Ford 2019).

Finally, two recent studies have examined the extent to which technology can be leveraged to increase access to college advising. One experiment conducted at a large Canadian university tested three treatment arms: one-on-one coaching, an online exercise, and a text-messaging support program. Only the one-on-one coaching arm showed significant results, potentially indicating limits in using electronic communication in helping foster longer-term academic performance (Oreopoulos and Petronijevic 2017). The other study evaluated a program at Georgia State University that leveraged artificial intelligence (AI) technology in developing a texting program with AdmitHub that sent customized messages to students guiding them through many aspects of the college enrollment process (Page and Gehlbach 2017). The “augmented intelligence technology” upon which the program was based made it possible for the computer to respond to a large majority of incoming questions, saving scarce time for college advisers and administrators. For the sample of students that had committed to attending Georgia State, the texting program increased enrollment there by 3.3 percentage points (Page and Gehlbach 2017).

6.3.3 Looking Forward

The studies reviewed in this subsection suggest that behavioral interventions aiding in the transition can exert substantial benefits. It would perhaps not have been intuitive to education economists a decade ago that a few text messages could make the difference between attending and not attending college, but recent evidence suggests precisely this possibility for a significant number of students. The studies on information treatments reviewed above, while far from definitive, suggest that videos might be a more effective communication medium for students (potentially) transitioning from high school to college than letters. Those studies relied on DVDs, but in the present day and age, texting or emailing videos to students is a more likely approach that could be tested further. Text messaging campaigns to remind students about financial aid and enrollment deadlines have shown consistent success; though over time, saturation of receiving text messages might dampen effects. Future research can further explore innovative ways to leverage university information systems to facilitate such campaigns, as well as varying the content, timing, and approach of messaging. More extensive application assistance is also an important strategy to follow, and it would be worth developing and testing computer programs—perhaps to be made available for free on websites—that help households confidently fill out FAFSAs. Last, while AI advising clearly has limits, future research can further explore which
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample Description</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
</table>
| Barr et al. (2016)      | EdPolicyWorks Working Paper  | Text messaging campaign prompting loan applicants at a large community college to make informed and active borrowing decisions | 1. Positive effect on reduction of unsubsidized loan borrowing  
2. Null impacts on enrollment, credit accumulation, and GPA | Main reported effect sizes (not standardized): 
1. Null on receipt of subsidized loan  
2. 3.1 percentage point decrease in share of students who received an unsubsidized loan and 5.3 percent decrease in amount of unsubsidized loans disbursed  
3. Null on enrollment, credit accumulation, and GPA | Community college loan applicants in Baltimore County | 2,807 students | Student               |
| Bergman, Denning, and Manoli (2019) | Journal of Policy Analysis and Management | E-mails and letters to potential/prospective/current college students on financial aid/incentives | Null effects | Null | Students who had applied to any public Texas college or university using the ApplyTexas.org portal. | 1,042,303 students | Student               |

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bettinger et al.</td>
<td><em>Quarterly Journal of Economics</em></td>
<td>FAFSA assistance during tax filing</td>
<td>1. Positive effect on FAFSA submissions and ultimately the likelihood of college attendance, persistence, and aid receipt for students in the combined assistance and information treatment</td>
<td>Main reported effect sizes (not standardized): 1. 15.7 percentage point increase in likelihood of filing FAFSA from being offered help to complete the form; null impact from information-only treatment 2. 8.1 percentage point increase in college enrollment among dependent students 3. 10.6 percentage point increase in Pell Grant receipt 4. 8 percentage point increase in persistence among dependent students</td>
<td>Individuals in families with incomes less than $45,000 and with a family member between the ages of 15 and 30 who did not have a bachelor's degree in Ohio and North Carolina</td>
<td>4,187</td>
<td>Individuals</td>
</tr>
<tr>
<td>Bird et al.</td>
<td>EdPolicyWorks Working Paper</td>
<td>Nudges for early FAFSA filing through the Common Application</td>
<td>Positive effect for treatment arm that involves concrete planning prompts</td>
<td>Main reported effect sizes (not standardized): 1. 1.1 percentage point increase in college enrollment</td>
<td>US high school seniors who had registered with the Common Application</td>
<td>454,243</td>
<td>Student</td>
</tr>
<tr>
<td>Castleman, Arnold, and Wartman</td>
<td><em>Journal of Research on Educational Effectiveness</em></td>
<td>Providing college counseling to low-income students during the summer through email, text message, and in-person consultation</td>
<td>Positive effects on both the rate and quality of college enrollment</td>
<td>Main reported effect sizes (not standardized): 14 percentage point increase in likelihood to enroll immediately in college and 19 percentage point increase in likelihood to keep the postsecondary plans they developed during senior year</td>
<td>Senior students across 7 high schools in Providence, Rhode Island</td>
<td>162 students</td>
<td>Student</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Castleman and Meyer (2016)</td>
<td>EdPolicyWorks Working Paper</td>
<td>A text messaging campaign to provide lower-income college students with simplified information, encouragement, and access to one-on-one advising</td>
<td>Positive effect on completion of more freshman year credits</td>
<td>Main reported effect sizes (not standardized): 2–2.23 additional credits over the year</td>
<td>Rising college students in West Virginia</td>
<td>1,198 students</td>
<td>Student</td>
</tr>
<tr>
<td>Castleman and Page (2015)</td>
<td>Journal of Economic Behavior &amp; Organization</td>
<td>Text messages to reduce summer melt among students with less access to college-planning supports and who were not as far along with their college planning at the completion of high school, null full sample impacts</td>
<td>Positive effect on enrollment</td>
<td>Main reported effect sizes (not standardized): 3.0 percentage point increase in 2-year college enrollment; Null on enrollment in 4-year colleges</td>
<td>Recent high school graduates in Dallas, Boston, Lawrence, Springfield, and Philadelphia</td>
<td>12,676 students</td>
<td>Student</td>
</tr>
<tr>
<td>Castleman and Page (2016)</td>
<td>Journal of Human Resources</td>
<td>Text message to improve FAFSA re-filing for sophomore year</td>
<td>Positive effects</td>
<td>Main reported effect sizes (not standardized): 14 percentage point increase in likelihood to remain continuously enrolled through spring of sophomore year among community college freshman; null impacts on sophomore year persistence among freshmen at 4-year institutions</td>
<td>First-time college freshmen in Massachusetts</td>
<td>808 students</td>
<td>Student</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castleman and Page (2017)</td>
<td>EdPolicyWorks Working Paper</td>
<td>Text messages to improve enrollment tasks</td>
<td>Positive effects (although no additional benefit from including parents on nudges)</td>
<td>Main reported effect sizes (not standardized): 3.1 percentage point increase in on-time college enrollment (although no additional benefit from including parents on nudges)</td>
<td>High school graduates in Boston, Fall River, Lawrence, and Springfield, Massachusetts; and Miami, Florida</td>
<td>4,754 students</td>
<td>Student</td>
</tr>
<tr>
<td>Chande et al. (2015)</td>
<td>Harvard Business School Working Paper</td>
<td>Texting motivational messages and organizational reminders to students, with messages drawing on insights from behavioral economics</td>
<td>1. Positive effects on reduction of the proportion of students that stop attending 2. Positive effects on increasing average attendance</td>
<td>Main reported effect sizes (not standardized): 36 percent decrease in the proportion of students that stop attending and a 7 percent increase in average attendance</td>
<td>Adult learners in England</td>
<td>1,179 students in 152 classes</td>
<td>Class</td>
</tr>
<tr>
<td>Darolia and Harper (2018)</td>
<td>Educational Evaluation and Policy Analysis</td>
<td>Letter e-mailed to students regarding financial aid</td>
<td>Null effects</td>
<td>Null</td>
<td>College students at the University of Missouri</td>
<td>9,802 students</td>
<td>Student</td>
</tr>
<tr>
<td>Hyman (2018)</td>
<td>University of Michigan Education Policy Initiative Working Paper</td>
<td>Mailing letters with web address to college information website</td>
<td>Null impact on college enrollment for full sample; positive impacts among poor students</td>
<td>Main reported effect sizes (not standardized): 1. Null 2. 1.4 percentage point increase in the probability that they will enroll in college</td>
<td>Eleventh grade public school students in Michigan</td>
<td>49,156 students</td>
<td>Student</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ksoll et al. (2014)</td>
<td>Center for Global Development Working Paper</td>
<td>Innovative mobile phone-based adult education program (Cell-Ed)</td>
<td>1. Positive effect on students' basic and broad reading scores 2. Positive effect on participants' self-esteem</td>
<td>Main reported effect sizes (not standardized): 1. 2–4 year increase in reading levels over a four-month period 2. 7 percent increase in participants' self-esteem</td>
<td>Adult learners in Los Angeles</td>
<td>70 students</td>
<td>Student</td>
</tr>
<tr>
<td>O'Connell and Lang (2018)</td>
<td>Journal of Research on Technology in Education</td>
<td>Personalized email reminders encouraging out-of-class study</td>
<td>Positive effects on exam performance</td>
<td>0.2 standard deviation increase on exam performance</td>
<td>First-year students at a mid-sized university in the northeast United States</td>
<td>281 students</td>
<td>Student</td>
</tr>
<tr>
<td>Oreopoulos and Dunn (2013)</td>
<td>Scandinavian Journal of Economics</td>
<td>3-minute video and opportunity to use financial aid calculator</td>
<td>Positive effects on understanding college-related benefit–cost</td>
<td>Main reported effect sizes (not standardized): 40 percent increase in ratio between expected earnings after college and after high school</td>
<td>Disadvantaged high school students in Canada</td>
<td>1,616 students</td>
<td>Student</td>
</tr>
<tr>
<td>Oreopoulos and Ford (2019)</td>
<td>Journal of Policy Analysis and Management</td>
<td>Application assistance incorporated into the high school curriculum. Program included a website that was designed to provide students a &quot;one-stop-shop&quot; with directed access to application websites, informational videos, tools for identifying suitable programs for each student, and a financial aid and budget calculator</td>
<td>Positive effect on application rates and college enrollment</td>
<td>Main reported effect sizes (not standardized): 14 percentage point increase in application rates and 5.2 percentage point increase in college enrollment with virtually all of this increase in 2-year community college programs</td>
<td>High schools in Canada with low college transition rates</td>
<td>11,356 students at 86 schools</td>
<td>School</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Interventions</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------------------</td>
<td>----------------------------------------</td>
<td>--------------------------------</td>
<td>------------------------------------------------</td>
<td>-------------------------------------------</td>
<td>-----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Oreopoulos and Petronijevic (2017)</td>
<td>NBER Working Paper</td>
<td>Text-based advising</td>
<td>Null effects found</td>
<td>Null</td>
<td>First-year college students in Canada</td>
<td>4,900 students</td>
<td>Student</td>
</tr>
<tr>
<td>Page, Castleman, and Meyer (2016)</td>
<td>SSRN</td>
<td>FAFSA texting program</td>
<td>Positive effect on enrollment among students with less access to college-planning supports</td>
<td>Main reported effect sizes (not standardized): 5–6 percentage point increase in FAFSA submission and completion; 4 percentage point increase in timely college enrollment</td>
<td>High schools in Texas and Delaware</td>
<td>17,000+ students in 66 high schools</td>
<td>Texas; school Delaware: not randomized—quasi-experimental</td>
</tr>
<tr>
<td>Smith et al. (2018)</td>
<td>Journal of Economic Education</td>
<td>Software that sends a “grade nudge,” a personalized message to each homework assignment regarding the student’s current grade. The message explains precisely how the assignment will impact the student’s final grade given their current standing in the class.</td>
<td>Positive effect on homework performance</td>
<td>3.3 percentage point increase in grades (with evidence suggesting a larger impact from nudges early in the semester)</td>
<td>First year microeconomics students at Washington State University</td>
<td>122 students</td>
<td>Student</td>
</tr>
</tbody>
</table>
tasks are adequately handled by AI systems and which are not.

6.4 Social-Psychology Interventions

6.4.1 Background, Context, and Mechanisms of Impact

Finally, several recent programs have been evaluated that take advantage of ed-tech to implement interventions aimed at providing students with social psychological support. It is widely agreed among education scholars that students’ educational performance is heavily affected by emotions and mentalities, which are in turn tied up in students’ perceptions of themselves in relation to those around them (Duckworth et al. 2007, Harackiewicz and Priniski 2018). The interventions discussed in this subsection center on short exercises administered to students that are intended to cultivate mentalities conducive to learning and thereby overcoming such barriers. Like the other behavioral interventions discussed in this section, social-psychology interventions provide subtle promotion of ideas that designers hope will alleviate psychological constraints on learning, at least among a subset of students susceptible to them. Within our theory-of-change framework, social-psychology interventions may help to improve the efficiency of education by cultivating noncognitive skills like self-efficacy and persistence. But whereas the other behavioral interventions discussed in this subsection aim to overcome cognitive limitations, such as procrastination or inattention, the present subsection’s interventions focus on emotions and beliefs, such as feeling out of place or “not smart.”

The class of social-psychology interventions that has been most prominent in scholarly discussions is that of “mindset interventions.” Academic mindsets consist of the “attitudes, beliefs, and dispositions about school and learning that are associated with positive academic outcomes and school success,” and some education theorists expect that they will shape educational outcomes by affecting “the quality, duration, and intensity with which students engage in critical academic behaviors … and deploy learning strategies” (Snipes et al. 2015). Typically, mindset interventions involve short reading and writing exercises that attempt to convince students that an individual’s intelligence develops over time and that new skills can be learned with practice (i.e., a “growth mindset”), rather than academic performance being an unchanging attribute (“fixed mindset”). Examples of other social-psychology interventions include programs that attempt to improve self- or values affirmation (Brady et al. 2016; Borman, Grigg, and Hanselman 2016), reduce stereotype threat (Cohen et al. 2006; Good, Aronson, and Inzlicht 2003; Hanselman et al. 2014), highlight relevance of school or academic work to one’s life goals (Hulleman and Harackiewicz 2009), and reinforce sense of purpose.

Social-psychology interventions can conceivably be delivered at any point during the educational life course, at least once an individual is old enough to have acquired the abilities for self-reflection and abstract thinking needed to make a mindset intervention’s material meaningful. The interventions tested so far have tended to cluster around the transition to high school and the transition to college. This is perhaps because it is easier to experimentally test samples of older youth in larger classrooms, or because it is typically understood that persistence in the face of difficulty becomes increasingly important in high school and college, when even students with strong academic and cognitive skills almost invariably run into material that challenges them and must decide how to interpret these difficulties. During transitional periods, as when entering high school or college, students may be more
open to outside influences, like the materials presented in a social-psychology intervention. By the time an individual has reached adulthood, these interventions may be less likely to deliver effects that are as strong, since the individuals’ thinking habits are likely to have crystallized more, and most of the academic work that could have benefited from improved mindset is likely behind the individual.

Several small-sample studies suggest that social-psychology interventions can yield significant learning effects, especially relative to the interventions’ typically low costs in time and money (Yeager and Walton 2011). However, other studies show minimal or null effects, particularly when larger sample sizes are used. For instance, while one of the pioneering mindset intervention studies showed an impact of 0.3 GPA points in a study involving 158 seventh graders from a Northeastern school, an attempt to replicate the study using the same mindset materials among 374 seventh graders from 11 schools in a single Midwestern school district found an impact of only 0.065 GPA points, and then an entirely null result when they replicated the experiment with 449 students (Cohen et al. 2006; Borman, Grigg, and Hanselman 2016; Yong 2016). An intervention targeting stereotype threat—when individuals are or feel themselves to be at risk of conforming to stereotypes about their social group—among minority and female students showed no significant effects in a sample of over 1,300 students in three urban high schools (Bancroft, Bratter, and Rowley 2017). A self-affirmation exercise, which encourages students to identify and reflect upon core personal values, showed mostly insignificant effects—and even a reduction in performance among girls—on a sample of 2,500 seventh–eighth graders from six middle schools in and around Philadelphia (Dee 2015). A recent meta-analysis found that average effects of these interventions are generally small or insignificant, but that there may be some signs of promise for certain subgroups, including students with low socioeconomic status (Sisk et al. 2018).

Another meta-analysis notes that experimental evaluations of social-psychology interventions in education typically find smaller effect sizes than do quasi-experimental evaluations (Lazowski and Hulleman 2016).

It is, as of yet, unclear the extent to which the mixed findings relating to social-psychology interventions reflect their general inefficacy at scale versus context-bound potential. Ed-tech offers a range of possibilities for testing and potentially improving these interventions. Up until the last few years, the typical model has been for students to complete program activities in classrooms using paper and pencils. Delivering interventions via computers may:

allow materials to be delivered to recipients exactly as designed without extensive researcher involvement or facilitator training … eliminate geographic constraints, opening access to students at multiple school sites and sites far from research centers; and …drastically reduce logistical burdens, the marginal cost of additional participants, and the costs of data collection and large-scale evaluation…. (Paunesku et al. 2015)

By offering to drastically reduce the costs of scale-up, delivering social-psychology interventions through computers can enable the wide-ranging and systematic testing of these interventions necessary to fully understand their workings, potential benefits, and limitations—in addition to reducing the cost side of the cost–benefit ratio for the treatment itself.

In particular, experimental research on ed-tech—although still in its infancy—has already begun to contribute toward scholarly understanding of education-oriented social-psychology interventions in three broad areas. First, evaluating the delivery through computers of traditional
social-psychology interventions previously tested using paper-and-pencil format allows researchers to evaluate these interventions on larger scales than would otherwise be possible. The endeavor also holds potential for generating evidence about the extent to which these interventions work differently when delivered through computers, as findings accumulate. Second, ed-tech research offers the potential for studying delivery of social-psychology interventions through media that is not possible in traditional paper-and-pencil approaches, like the use of audiovisual content and computer games. Third, ed-tech can facilitate iterative experimentation and rapid feedback for learning what works best between alternative intervention variations. Many colleges and high schools could test social-psychology interventions regularly and converge on what helps most, for whom, and when. The next subsection reviews the evidence that has accumulated so far on each of these topics.

6.4.2 Experimental Evidence

We identified 15 recent experimental studies that have tested the extent to which traditional social-psychology interventions are effective when computers are used for delivery. Of these, two were carried out on small to medium scales comparable to those of other studies from this literature that do not use ed-tech. Findings are similar, and the authors of these studies conclude that the interventions appear promising. In one of these studies (Morisano et al. 2010), the authors test the effects of an online written 2.5-hour goal-setting intervention on a sample of 85 first-year McGill University students, finding a positive impact of 0.5 SDs on end-of-semester GPA, as well as positive and significant impacts on full-time enrollment and self-reported affect. However, the study’s sample size is very small and was selected from volunteers who responded to an advertisement—the students who might be most likely to enthusiastically make use of such an intervention—making it difficult to assess the external validity of these findings. Indeed, an attempt to replicate this online goal-setting exercise by one of this article’s coauthors (Oreopoulos) found precise null effects for a sample of nearly 2,000 first-year university students, as well as for subgroups of students more at risk of poor academic performance (Oreopoulos, Patterson, et al. 2018).

Similarly, a second study (Yeager, Henderson, et al. 2014) tests a web-based “prosocial,” “purpose-for-learning” intervention—which teaches students to learn with the goal of making a broader impact—on 338 ninth graders in a relatively affluent and high-performing San Francisco Bay Area high school, and finds significant impact on STEM GPA (\(d = 0.21\))—but only for students with baseline GPAs below 3.0 out of 4.0. Pooled-sample impacts are not statistically significant. While in some ways, significant effects in a well-off school like the one studied here may represent a strong test, since more privileged students are more likely to have other forms of social psychological support, guidance from teachers and use of other resources in a decent school environment may be a necessary link enabling students to channel improved motivations into higher GPA, making generalizability of these findings to struggling schools difficult to judge. In fact, poorly performing students in high-achieving schools—one way to describe the subsample of this study that benefited from the program—may well be an especially likely group to obtain positive impacts, since they face some challenges and have scope for improvement, but also have the supports and resources to succeed should they choose to increase their efforts. Furthermore, the outcome examined is very short term: after only one quarter of the school year. Few studies, including this one, test for longer-lasting effects.
While educational psychologists have been intrigued by social-psychology interventions since the early 1990s, researchers have only recently begun testing efforts to scale up in ways that explicitly account for the changes that are likely to come with the transition from small, controlled settings to real-life policy situations. Two recent studies present stronger evidence on social-psychology interventions than the two just discussed, given their larger and more representative samples, rigorous methods, and scaled-up implementation structures. Both of these studies reinforce the proposition—perhaps more strongly than before, given their large and diverse sample sizes and rigorous methods—that social-psychology interventions can, under some circumstances, lead to meaningful effects, but that these effects tend to be concentrated within subsamples and, even then, tend to be quite small (not to mention that whether these effects can last beyond a single semester or school year remains an open question).

The first large-scale technology-based mindset intervention study was conducted in 2015 (Paunesku et al. 2015), testing the effects of two interventions—one “growth mindset” intervention, and one “sense-of-purpose” intervention—on academic performance. The sample included 1,594 students from 13 different high schools. The authors purposely included a diverse sample of schools in order to maximize external validity, despite the accompanying loss in statistical power, choosing schools that varied across geography, socioeconomic characteristics, and public versus charter versus private status. The intervention consisted of two 45-minute computer sessions—the first a growth mindset module or control, and the second a sense-of-purpose module or control—separated by roughly two weeks. The sense-of-purpose intervention sought to cultivate a sense of prosocial purpose for studying hard, while the control materials highlighted economic self-interest. The study found positive and statistically significant—although substantively small—impacts on average grades for core courses, for the bottom third of students, of 6.4 percentage points for both of the interventions, although there was no added benefit to receiving both treatments over receiving just one. No significant effects are reported for the full sample.

The second large-scale study (Yeager, Walton, et al. 2016) tested several variations of a mindset intervention given to incoming college students in advance of matriculation. The researchers ran three experiments: one on 584 graduating seniors from four urban charter schools going on to a variety of two- and four-year colleges; one on 7,335 incoming students at a high-quality public four-year university; and one on 1,592 incoming students at a private university with competitive admissions. Using pooled effects from the different mindset variations, experiment 1 saw an improvement of full-time continuous enrollment in the first year of 9 percentage points (from 32 percent to 41 percent); experiment 2 showed significant effects on disadvantaged students’ enrollment (as defined by racial minority and/or first-generation college attendance status—which accounted for all participants in experiment 1), but not advantaged ones, leading to a drop in the gap between disadvantaged and advantaged students in first-year full-time continuous enrollment by 40 percent (the rate increased from 69 percent to 73 percent in the disadvantaged group and stayed at 79 percent in the advantaged group). Experiment 3 found a positive impact on GPA of 0.25 standard deviations for disadvantaged students. While the interventions overall seem to have left an impression on the students, only the third experiment reported effects on learning outcomes, and here significance was again restricted to a subsample. Furthermore, the study used a factorial design to test the
separate and joint impacts of two different social-psychology interventions, mindset and social belonging, and only the latter leads to any significant impact when separated out. Given the large-scale and diverse sample of the study, the reported impacts raise serious questions about the generalizability of ostensibly positive effects from mindset interventions found in small, highly controlled samples, while lending new evidence to the potential promise of social-belonging interventions.

Two other experimental studies show that social-psychology interventions can exert psychological effects, but without necessarily improving final outcomes. One of these studies (Unkovic, Sen, and Quinn 2016) found that encouraging e-mails can raise the probability that graduate students will apply to a competitive conference, with stronger effects for women than for men in a gender-imbalanced field. However, these effects did not translate into higher acceptance rates, and women in the control group ended up with higher acceptance rates than those in the treatment groups. In this case, the intervention probably did not cause harm to the female graduate students with unsuccessful applications, and taking academic risks is generally considered a positive step. But the intervention would likely have needed to blend in additional resources in order to be effective in increasing final outcomes, a lesson that may prove to be more general. A small-sample study (Forsyth et al. 2007) conducted within a single undergraduate psychology class even found that an intervention aimed at bolstering self-esteem significantly decreased final exam scores for students already receiving Ds and Fs, suggesting that overconfidence could be an unintended consequence of self-esteem interventions that lack additional components.

In addition to more efficiently delivering interventions that are similar or identical in content to paper-and-pencil interventions, ed-tech may also add to social-psychology interventions by facilitating intervention models that use technology-enabled multimedia going beyond simple text. Although literature testing such studies is still in its infancy, three recent studies provide some initial insight into this area. One study (Walton et al. 2015) tested an intervention on a sample of 228 first-year students at the University of Waterloo that was delivered in a face-to-face classroom setting, but involved participants listening to audio testimonies from fellow students to reinforce mindset messages. The experiment found large positive effects of more than a standard deviation on semester GPA for women (no effects for men)—although only within male-dominated majors (i.e., >80 percent male). This large effect size suggests the potential value of future tests of related intervention models that use multimedia rather than traditional reading and writing, although the present study provides limited evidence on its own given the narrow sample size, short time frame, and restriction to male-dominated majors.

Two other education-focused social-psychology studies (Yeager, Trzesniewski, and Dweck 2013; Yeager, Johnson, et al. 2014) involve a computer game known as Cyberball that engages the participant in an electronic game of catch with two other individuals that are allegedly other students from the same school. The game can create a sense of exclusion if the other “players” begin by aiming the ball at the participant but then begin ignoring the participant—this then allows researchers to test variation in students’ reactions to exclusion. The game functions as a measurement device within these two studies, rather than part of the intervention itself, but the studies nonetheless help to highlight the potential role that computer games can play in testing social-psychology interventions, and one can imagine potential applications through which games become part of the intervention.
Finally, one study (Yeager, Romero, et al. 2016) shows the potential utility of ed-tech as a way to iteratively improve the design of mindset interventions. Yeager and colleagues take the growth mindset intervention used by Paunesku et al. (2015) and use ICT in an attempt to improve the intervention using techniques of “design thinking.” In particular, they modified the intervention content using qualitative pretesting in combination with quantitative A/B testing using Amazon Mechanical Turk with about 3,000 participants. The improvement process led to several changes, including the incorporation of quotes from celebrities and the use of bullet points instead of paragraphs. The researchers then tested the revised intervention in two studies, one with a sample of 7,501 ninth graders across 69 high schools in the United States and Canada (looking at proximate outcomes, i.e., the impact of the intervention on growth mindset), and one with a sample of 3,676 ninth graders representing 95 percent of the students at 10 schools. Both studies found positive and significant effects when comparing the revised version of the intervention to the original (there was no pure control group entirely without a mindset intervention, and thus the study says nothing about the impact of the mindset intervention itself). Effect sizes on the main outcome of interest—an index for challenge-seeking behavior—were small, typically below 0.10 standard deviations, although larger for lower-performing students than for average or higher-performing ones. But these impacts are still significant considering that the control group also received a very similar mindset intervention.

6.4.3 Looking Forward

Given its potential to reduce costs, incorporate multimedia, and identify treatment and population characteristics that yield the largest effects, ed-tech represents a promising route for the study and honing of social-psychology interventions. This does not mean that ed-tech can, on its own, make these interventions effective or worthwhile. The studies by Paunesku et al. (2015) and Yeager, Romero, et al. (2016) both contained large and diverse samples and used rigorous methods—if there were large and wide-ranging effects of the social-psychology interventions on academic outcomes, these studies would likely have caught them. Instead, the studies reinforce a picture in which social-psychology interventions can have significant and meaningful effects relative to their low costs, but that these effects tend to be small, and are statistically significant only for subsamples.

Future ed-tech research can advance scholarly understanding of social-psychology interventions in several directions. First, future studies should attempt to refine understanding of the specific student types that are most likely to significantly gain from these interventions. Enough studies have been done to demonstrate that there is no meaningful answer as to whether social-psychology interventions “work” or not—they work differently and to different extents among different populations (and, presumably, depending on specific design features). Whether researchers and practitioners can efficiently identify the conditions under which these interventions are likely to work can make all the difference as to whether these interventions are cost-effective relative to the next best option.

Findings from studies so far have generated hints in this direction. For instance, effect sizes tend to be larger for those who start out further behind in terms of academic performance and/or social-psychological attitudes. However, the evidence they provide is far from sufficient in this regard because these studies were typically designed to measure full-sample, rather than subgroup, effects. Even where studies are adequately powered to test heterogeneity, contemporary best
practices in experimental research demand detailed preliminary plans for heterogeneity tests and, wherever possible, documented pre-analysis plans. A team of researchers is currently leading exactly such an endeavor—a study aiming specifically to identify heterogeneity with a detailed analysis plan on a nationally representative sample of 20,000 ninth graders. These researchers are considering variation not only in student characteristics, but also in school and classroom conditions and resources. Future research should pursue a similar agenda in a college setting and should involve new research teams (since researchers from Stanford University and collaborators have conducted nearly all of the high-quality mindset experiments in the ed-tech space).

Second, more studies are needed that directly compare different social-psychology intervention models—for instance, mindset versus sense-of-purpose versus stereotype threat reduction—within the same experimental context, so that their significance and effect sizes may be compared to one another. Different social-psychology interventions may target different problems and hold different theories of change—there is no reason that the same findings should hold for different intervention types. Furthermore, alternative models may work differently on different populations.

Third, the literature on ed tech-enabled multimedia social-psychology interventions has, perhaps surprisingly, barely begun. The Walton et al. (2015) study described above uses only audio technology; perhaps varying blends of audio, video, and text could produce emotional responses capable of engendering lasting change in students. Furthermore, games have thus far been used for testing rather than intervention purposes. Psychologists have emphasized the importance of participants being engaged, which is why they are often asked to write essays or letters to hypothetical future students rather than simply absorbing the information. But games may offer more active forms of engagement than writing prompts, particularly for younger children.

Fourth, while several recent collaborators have already spearheaded the use of feedback techniques in mindset interventions, more work in this direction is needed, given that this is the only study that has incorporated ICT-based feedback and that the A/B testing used adults on Mechanical Turk rather than the population of interest itself. This means that the quality of customization may have been limited, and thus that the effects may represent a lower range. Furthermore, qualitative piloting—which should be a requirement for almost any program evaluation regardless—was used to alter the treatment relative to control in addition to ICT-based feedback, making it difficult to disentangle the impacts. Beyond developing interventions that are more effective for larger groups of students, feedback could be used to customize interventions for students with different personality types, or socioeconomic or identity backgrounds, leading to improved targeting.

Finally, while the focus so far on the transitions to high school and college is understandable, future research should test social-psychology interventions delivered through ed-tech within different age ranges. While it may be unlikely that young children would be affected by these interventions, middle school children might benefit from them, and such early interventions could yield especially large effects to the extent that these effects are recursive, as discussed above. Furthermore, younger children might be more likely to respond well to engaging multimedia content on a computer versus detailed articles and writing prompts. And,
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forsyth et al. (2007)</td>
<td>Journal of Social and Clinical Psychology</td>
<td>Self-esteem bolstering intervention</td>
<td>1. Negative results on self-esteem for the D and F students 2. Null impacts for students in the other conditions.</td>
<td>Main reported effect sizes (not standardized): the D and F students got worse as a result of self-esteem bolstering and students in the other conditions did not change (exact magnitude not reported)</td>
<td>US college students</td>
<td>90 students</td>
<td>Student</td>
</tr>
<tr>
<td>Good, Aronson, and Inzlicht (2003)</td>
<td>Applied Developmental Psychology</td>
<td>E-mail mentorship by college students who encouraged middle school students to view intelligence as malleable or to attribute academic difficulties in the seventh grade to the novelty of the educational setting</td>
<td>1. Positive effect on math standardized test scores for females 2. Positive effect on reading standardized test scores</td>
<td>1. Math scores for female students: 1.13 standard deviations (incremental condition); 1.50 standard deviations (attribution condition) 1.30 standard deviations (combined condition) 2. Reading scores: 0.52 standard deviations (incremental condition) and 0.71 standard deviations (attribution condition)</td>
<td>Seventh grade students in Texas</td>
<td>138 students</td>
<td>Student</td>
</tr>
<tr>
<td>Harackiewicz et al. (2012)</td>
<td>Psychological Science</td>
<td>3-part intervention (2 brochures mailed to parents and a website) highlighting the usefulness of STEM courses</td>
<td>Positive effect on taking nearly 1 semester more of science and mathematics in the last 2 years of high school</td>
<td>Main reported effect sizes (not standardized): nearly an extra semester of math and science courses in the last 2 years of high school</td>
<td>High school students in Wisconsin</td>
<td>188 students and their parents</td>
<td>Students and parents</td>
</tr>
<tr>
<td>Morisano et al. (2010)</td>
<td>Journal of Applied Psychology</td>
<td>Goal-setting program</td>
<td>Positive effects on academic performance compared with the control group after 4-month period</td>
<td>1. 0.50 standard deviation increase in GPA 2. 0.46 decrease in negative affect scores 3. Null for enthusiasm scores</td>
<td>College students in Canada</td>
<td>85 students</td>
<td>Student</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>--------------------------------------------------</td>
<td>-------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Oreopoulos, Patterson, et al. (2018)</td>
<td>NBER Working Paper</td>
<td>Online planning exercise with information and guidance to create a weekly schedule containing sufficient study time and other obligations</td>
<td>Null impacts on course grades, credit accumulation, or retention</td>
<td>Null</td>
<td>1. Selective urban college</td>
<td>9,000+</td>
<td>Student</td>
</tr>
<tr>
<td>Oreopoulos, Petronijevic, et al. (2018)</td>
<td>NBER Working Paper</td>
<td>Choose-Your-Own-Challenge online modules designed to teach students effective learning behaviors and adaptive perspectives</td>
<td>Null impacts on course grades, GPA, or credit accumulation</td>
<td>Null</td>
<td>First-year university students at a Canadian university</td>
<td>3,395</td>
<td>Student</td>
</tr>
<tr>
<td>Paunesku et al. (2015)</td>
<td>Psychological Science</td>
<td>Growth mindset and sense-of-purpose interventions</td>
<td>Positive effect on GPA in core academic courses and increased the rate at which students performed satisfactorily in core courses among students at risk of dropping out of high school</td>
<td>Main reported effect sizes (not standardized): 6.4 percentage point increase in satisfactory performance in core courses</td>
<td>US students in 13 geographically diverse high schools</td>
<td>1,594</td>
<td>Student</td>
</tr>
<tr>
<td>Rege et al. (forthcoming)</td>
<td>American Psychologist</td>
<td>A program teaching a growth mindset of intelligence</td>
<td>Positive program effects on challenge-seeking behavior with modest heterogeneity</td>
<td>0.23 decrease in number of “easy” problems taken.</td>
<td>US ninth grade students</td>
<td>14,866</td>
<td>Student</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Author</th>
<th>Publication</th>
<th>Intervention</th>
<th>Direction of effect</th>
<th>Effect size</th>
<th>Sample</th>
<th>Sample size</th>
<th>Unit of randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unkovic, Sen, and Quinn (2016)</td>
<td>PLOS One</td>
<td>Personalized emails encouraging graduate students to apply for a conference</td>
<td>Positive effect among application rates; null results on acceptance rates to conference</td>
<td>Main reported effect sizes (not standardized): 1. 2.7 percentage point increase in applications (larger effect for women) 2. Null for conference acceptance</td>
<td>US graduate students</td>
<td>3,945 students</td>
<td>Student</td>
</tr>
<tr>
<td>Walton et al. (2015)</td>
<td>Journal of Educational Psychology</td>
<td>1. Social-belonging intervention to protect students’ sense of self-belonging 2. Affirmation-training intervention to help students manages stress related with social marginalization</td>
<td>Positive effects on women's school-reported engineering GPA</td>
<td>1.04 standard deviation increase in women's GPA; null impacts for men</td>
<td>First-year students at the University of Waterloo</td>
<td>228 students</td>
<td>Student</td>
</tr>
<tr>
<td>Yeager et al. (2013)</td>
<td>Child Development</td>
<td>6-session intervention that taught an incremental theory (a belief in the potential for personal change) through Cyberball electronic game.</td>
<td>Positive effects on reducing aggressive behavior and increasing prosocial behavior among the incremental theory group</td>
<td>Main reported effect sizes (not standardized): Nearly 40 percent reduction in aggressive retaliation after a controlled provocation (an experience of exclusion), 300 percent increase in prosocial behavior</td>
<td>Ninth and tenth grade students in California</td>
<td>230 students</td>
<td>Student</td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>-------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Yeager, Henderson, et</td>
<td><em>Journal of Personality and Social</em></td>
<td>Promoting a prosocial, self-transcendent purpose</td>
<td>Positive effects on persistence on a boring task and persistence in college.</td>
<td>Main reported effect sizes (not standardized): double the amount of time students spent on tedious exam review questions; 35 percent increase in the number of boring math problems</td>
<td>US high school and college students</td>
<td>1,364 students</td>
<td>Student</td>
</tr>
<tr>
<td>et al. (2014)</td>
<td>Psychology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Study 1:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,364 students</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Study 2:</td>
<td>338 students</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Study 3:</td>
<td>89 students</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Study 4:</td>
<td>429 college students</td>
<td></td>
</tr>
<tr>
<td>Yeager, Johnson, et</td>
<td><em>Journal of Personality and Social</em></td>
<td>A malleable (incremental) theory of personality—the belief that people can change.</td>
<td>Positive effects on reactions to an immediate experience of social adversity and better academic performance; lower overall stress and physical illness.</td>
<td>Main reported effect sizes (not standardized): the incremental theory group showed less negative reactions to an immediate experience of social adversity and, 8 months later, reported lower overall stress and physical illness. They also achieved better academic performance over the year.</td>
<td>Ninth grade students in California</td>
<td>158 students</td>
<td>Student</td>
</tr>
<tr>
<td>et al. (2014)</td>
<td>Psychology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Publication</td>
<td>Intervention</td>
<td>Direction of effect</td>
<td>Effect size</td>
<td>Sample</td>
<td>Sample size</td>
<td>Unit of randomization</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
<td>-------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Yeager, Romero, et al. (2016)</td>
<td><em>Journal of Educational Psychology</em></td>
<td>Growth mindset interventions during the transition to high school: qualitative inquiry and rapid, iterative, randomized “A/B” experiments were conducted to inform intervention revisions for this population.</td>
<td>Positive effect on ninth grade core-course GPA and reduced D/F GPAs for lower achieving students when delivered via the Internet</td>
<td>Main reported effect sizes (not standardized); the intervention was an improvement over previous versions in terms of short-term proxy outcomes and it improved ninth grade core-course GPA and reduced D/F GPAs for lower achieving students when delivered via the Internet</td>
<td>US ninth grade students</td>
<td>Study 1: 7,501 students</td>
<td>Student</td>
</tr>
<tr>
<td>Yeager, Walton, et al. (2016)</td>
<td><em>Proceedings of the National Academy of Sciences</em></td>
<td>“Lay theory” intervention that explains the meaning of commonplace difficulties before college matriculation</td>
<td>Positive effect on full-time enrollment rates and grade point averages and reduced the overrepresentation of socially disadvantaged students among the bottom 20 percent of class rank</td>
<td>Main reported effect sizes (not standardized); 9 percentage point increase in full-time enrollment (lay theory intervention—experiment 1); null (growth mindset intervention—experiment 1); 4 percentage point increase in full-time enrollment among disadvantaged students (lay theory intervention—experiment 2); null for advantaged students (lay theory intervention—experiment 2); 0.09 increase in GPA for disadvantaged students for disadvantaged students (experiment 3)</td>
<td>US high school and college students</td>
<td>Study 1: 584 students</td>
<td>Study</td>
</tr>
</tbody>
</table>
although mindsets may crystallize later in life, interventions still could have effects for nontraditional students considering pursuing education or training later in life.

7. Conclusion

Technology has transformed large segments of society in ways that were once considered unimaginable. Education is no exception. Around the world, there is tremendous interest in leveraging technology to transform how students learn. In the coming years, new uses of ed-tech will continue to flood the market, providing students, parents, and educators with a seemingly limitless array of options. And experimental literatures are beginning to emerge in new domains, including in-class technology like iClickers (Lantz and Stawiski 2014) and adult education offered through text messages and other new platforms (Ksoll et al. 2014).

Amid the buzz and sizeable investment in ed-tech, we aim to step back and take stock of what we currently know from the experimental evidence in this nascent field and shed light into how technology may impact the education production function. This review hopes to advance the knowledge base by identifying and discussing the most promising uses of ed-tech to date that have been explored with experimental research and highlighting areas that merit further exploration. We contextualize our discussion within a framework for how education technology may or may not address traditional education constraints and increase the efficiency of investments in cognitive and noncognitive skill formation. Our review categorizes the existing literature into four categories: (i) access to technology, (ii) CAL, (iii) online courses, and (iv) technology-enabled behavioral interventions. Taken together, these studies and their potential mechanisms of impact suggest a few areas of promise for ed-tech and potential reasons for why ed-tech in some circumstances successfully contributes to the development of cognitive and noncognitive skills.

We found that simply providing students with access to technology yields largely mixed results. At the K–12 level, much of the experimental evidence suggests that giving a child a computer may have limited impacts on learning outcomes, but generally improves computer proficiency and other cognitive outcomes. While access to technology likely improves the learning environment by expanding opportunities for learning, by and large, increased access alone does not seem to advance cognitive skill formation. One bright spot that warrants further study is the provision of technology to students at the postsecondary level, an area where the limited RCT evidence has reported positive impacts. This may be due to the increased necessity of technology—and perhaps technology’s more critical role in skill formation—during later stages of the education life cycle.

From our review, CAL and technology-enabled behavioral interventions emerge as two areas that show considerable promise. CAL has the potential to advance cognitive skill formation by mitigating standard teacher and classroom constraints, thereby increasing the efficiency of investments. Especially when equipped with a feature of personalization, CAL has been shown to be quite effective in helping students, particularly with math. Two interventions in the United States stand out as being particularly promising—a fairly low-intensity online program that provides students with immediate feedback on math homework was found to have an effect size of 0.18 standard deviations, and a more intensive software-based math curriculum intervention improved seventh and eighth grade math scores by a remarkable 0.63 and 0.56 standard deviations. These results mirror those from promising interventions
examined in the developing country literature, such as the adaptive Indian learning software Mindspark, which was found to have large, positive impacts on math and Hindi. The consistent results of personalized software across different contexts suggest that this feature may be an effective mechanism that helps to overcome common educational challenges in classrooms with heterogeneous learning levels. However, it should be noted that quality of implementation is an important feature of CAL, and the degree to which a software successfully helps teachers to engage students in dedicated time spent on academic learning that appropriately matches student’s actual learning levels may be an important determinant of CAL software’s ability to drastically improve student learning. Far more research is needed to help us isolate the mechanisms for when and why certain CAL programs improve cognitive skill formation and where these findings may be generalizable.

Like with CAL, evaluations of behavioral interventions generally find positive effects across all stages of the education life cycle, although they are generally smaller than those found with the most effective CAL models. Given that technology-enabled behavioral interventions generally target noncognitive skill formation, it is perhaps unsurprising that the positive impacts found in this category were, for the most part, smaller in magnitude than CAL models that successfully augmented the development of cognitive skills that may be crucial to understanding specific concepts. At the same time, technology-enabled behavioral interventions, such as large-scale text message campaigns, are often incredibly cheap to carry out and hold great promise as a cost-effective solution to many challenges associated with behavioral barriers in education. Given the promising impacts reported across all stages of the education life cycle, it would be valuable to build on and test technology-enabled models that can further facilitate the development of noncognitive skills in a cost-effective, scalable manner.

Though online learning courses have exploded in popularity over the last decade, there continues to be limited rigorous research on its effectiveness. From our review, we found that online courses, in some cases, do increase access to education, particularly when easy alternates are not available. Yet relative to courses with some degree of face-to-face teaching, students taking online-only courses may experience negative learning outcomes. This may be due to the difficulty of online courses to replicate the pedagogical strengths of in-person instruction, including noncognitive characteristics such as the rapport between a teacher and student. On the other hand, the effects of blended learning are generally on par with those of fully in-person courses. This suggests that the appropriate combination of online and in-person learning may be cost effective, but further research on cost effectiveness of online learning are needed to better understand its potential.

Nevertheless, we also note the potential unintended consequences of ed-tech, and encourage more work to identify these mechanisms for both failure and success with more clarity through further research. The evidence on access to technology and online learning seems to caution against merely providing technological devices or transferring content to online platforms for technology's sake, which suggests that merely improving the learning environment with technology is insufficient if the aim is to bolster learning more broadly. On the other hand, a body of positive evidence exists for using technology to overcoming binding constraints to skill formation, whether that be through noncognitive or cognitive processes. However, it should be noted that technology without a mechanism for facilitating skill formation may serve as
a distraction with negative consequences. There is empirical evidence that concerns about “brain drain”—the hypothesis that the mere presence of a phone may undercut cognitive performance—may be substantiated. For example, one experimental study found that even when people are successful at not checking their phones, the mere presence of a smartphone reduces cognitive capacity, and that these costs are highest for those with the greatest smartphone dependence (Ward et al. 2017).

Given that our review specifically focuses on those ed-tech products and strategies that have been evaluated using experimental methodologies, it is important to note that there are many popular ed-tech products that have not been evaluated using experimental methods to date, and that the ones that have been tested may be limited in their generalizability given recruitment and study sample characteristics. This, along with the limited scope of experimental research, even on products and strategies that show promise, suggests that there is still a substantial need for more rigorous research on ed-tech. We recommend the following areas for future experimental research in ed-tech.

First, replications of interventions showing promise, such as CAL and technology-enabled behavioral nudges, are needed to better understand both the generalizability and the credibility of the current findings. We recognize that the large effect sizes in some of these studies may be cause for skepticism, but the fact that such effects emerge in more than one study by different authors suggests that these models may truly be effective and worth exploring in greater depth. Such results warrant replication to see if the same effect is found in different contexts and with different populations of students, and over longer periods of time.

For CAL, additional research is needed to understand to what extent the observed impacts are related to specific implementation models. Open questions also remain regarding the underlying mechanisms of effective CAL programs, specifically how the software interacts with teachers and current curriculum. Additionally, it is worth noting that many CAL interventions that are currently being used by schools and families have not been rigorously tested, and therefore the impacts of such CAL products that do not retain the characteristics we identify as promising mechanisms of impact—personalization, adaptivity, or fast feedback of data—remain unknown. We caution policy makers and educational researchers to not take labels of CAL software at face value, but to discern the potential mechanisms of impact before judging the promise of a particular software. For example, to what extent does a software that claims personalization truly adapt to match lessons to a student’s learning level without requiring the teacher to recognize and redirect struggling students? Additionally, as mentioned previously, the personalized and adaptive component of CAL software seems to be more consistently effective for math, yet a handful of studies show that CAL can be effective for language as well. More research will be needed to test mechanisms and impacts of newly emerging CAL products of varying subjects.

Given the promising evidence on technology-enabled behavioral interventions, researchers should prioritize understanding when these nudges most effectively support noncognitive skill formation by varying the timing and content of messages and testing how these models interact with other educational supports. New experimental studies should test natural extensions of existing behavioral models, for example messaging campaigns with a high degree of personalization or parent-information flow systems that draw on predictive analytics to trigger timely alerts to parents based on students’ performance in school. As artificial intelligence enters the realm of education technology,
more research is needed to understand the extent to which AI systems can substitute for human guidance in navigating complex processes, such as helping students fill out the FAFSA or providing parents with actionable guidance on how to help a student who is at risk of dropping out of school. Meanwhile, subsequent studies may need to evaluate different methods of delivery of messages such as social media. With many parents, teachers, and students being inundated with text messages on a daily basis, engagement with these messages may decrease with time. Taken together, the experimental research on technology-enabled behavioral interventions is quite promising, and additional research should focus on building upon this already exciting body of evidence.

More work to understand effects of online learning is also needed. We recognize that the rigorous evidence is sparse, but given its growing popularity, this is all the more reason to recognize what studies do exist and pursue further work. Future research can help us better understand costs and cost effectiveness of online learning, and to test different types of online and blended courses relative to face-to-face. In particular, there is a need for studies disentangling why online learning on its own does worse than face-to-face in some circumstances. Is it the loss of interaction with instructors and peers, or is it the lack of a structured environment for time management and study skills, or both? Better understanding these issues will provide information on whether it is possible to make online learning work better if a blended component is not possible, and if so, how. Additionally, as the online learning field is constantly evolving, new research is needed to understand how new models—such as MicroMasters programs and nanocredentials—may impact or democratize learning.

Fourth, more experiments that identify the key mechanisms of impact at different stages in the process of skill formation, and that measure longer term outcomes, are also needed. We need better theories to model how and when technology is likely to help processes of skill formation so that we can design studies that specifically test these hypotheses. We hope that the discussion in this paper will provide some roadmaps to testing potential mechanisms of impact so that findings can be more generalizable, and that designing rigorous studies that measure long-term outcomes is more feasible. We recognize that there is a shortage of RCTs and encourage more research that follows samples over time to determine whether effects fade or sustain in the long term. Given the framework described in this paper, there is also a need to better understand how ed-tech impacts children of younger ages. The role of technology to facilitate skill formation at these younger ages is likely to be different, so as more and more technologies are developed for children of all ages, these potential mechanisms also need to be tested.

Finally, more rigorous evaluations of popular ed-tech products that have not yet been evaluated experimentally. Evaluations on products such as flipped classrooms, smart boards, or even virtual reality technologies to explore other parts of the world or to foster empathy are also needed. We recognize that some types of products are less amenable to an experimental design than others, but this challenge in itself should be a part of the conversation about which ed-tech products and strategies are working and the extent to which we have certainty about their impacts. Other popular products, such as Khan Academy, have begun engaging in rigorous evaluations over the last few years, but studies are not yet complete or results are not yet published. Researchers should keep their eye on this emerging work and continue to build on this.

The ed-tech field is rapidly changing, and innovative tools and programs are frequently considered out-of-date after only several years. When faced with purchasing
decisions, education administrators often demand research that is timely, relevant, and actionable. The direction and form of the research may need to change to integrate more seamlessly into decision making. New tools have emerged to address some of these challenges, including Mathematica’s Ed-Tech Rapid-Cycle Evaluation Coach and the EduStar RCT platform. While rapid-cycle product testing is of course valuable, more research is needed to evaluate how underlying mechanisms—rather than a specific product—can advance learning. In the end, it should not be about the most popular product or even necessarily the technology itself, but about the best way to help students of all ages and levels learn.

References


Balu, Rekha, Kristin Porter, and Brad Gunton. 2016. Can Informing Parents Help High School Students Show up for School? Results from a Partnership between New Visions for Public Schools and MDRC. Oakland: MDRC.


Beal, Carole, Shandy Hauk, Steven A. Schneider, Weiling Li, and Christopher Harrison. 2013. “Randomized Controlled Trial (RCT) Evaluation of a Tutoring System for Algebra
Readiness.” https://pdfs.semanticscholar.org/0742/747ebc378e2ace1e94e6beh0fd1be4ad91e.pdf?_ga=2.37305632.1017916876.1565637357-1422954928.1537456066.


Davis, Dan, René F. Kizilcec, Claudia Hauff, and Geert-Jan Houben. 2018. “Scaling Effective Learning Strategies: Retrieval Practice and Long-Term


Experiment.” *Active Learning in Higher Education* 17 (1): 39–49.


He, Fang, Leigh L. Linden, and Margaret MacLeod. 2007. “Helping Teach What Teachers Don’t Know: An Assessment of the Pratham English Language Program.” https://pdfs.semanticscholar.org/9192/418b948e7447d0013e44c9a06f345b80142.pdf.


Jackson, Kirabo, and Alexey Makarin. 2018. “Can Online Off-the-Shelf Lessons Improve Student


North-Holland.


Rodríguez, C., Fabio Torres, and Juliana Zúñiga. 2011. “Impact of the ‘Computers to Educate’ Program on student dropout, school achievement and income to higher education.” CEDE Documents ISSN 1657-7191.

Rogers, Todd, and Avi Feller. 2018. “Reducing Student Absences at Scale by Targeting Parents’ Misbeliefs.”
Escueta et al.: Upgrading Education with Technology


