

Two Stances, Three Genres, and Four Intractable Dilemmas for the Future of Learning at Scale

Justin Reich

Massachusetts Institute of Technology
Cambridge, MA
jreich@mit.edu

ABSTRACT

The late 2000s and 2010s saw the full arc of a dramatic hype cycle in learning at scale, where charismatic technologists made bold and ultimately unfounded predictions about how technologies would disrupt schooling systems. Looking toward the 2020s, a more productive approach to learning at scale is the tinkerer's stance, one that emphasizes incremental improvements on the long history of learning at scale. This article offers two organizational constructs for navigating and building on that history. Classifying learning-at-scale technologies into three genres—*instructor-guided*, *algorithm-guided*, and *peer-guided* approaches—helps identify how emerging technologies build on prior efforts and throws into relief that which is genuinely new. Four as-yet intractable dilemmas—the *curse of the familiar*, the *edtech Matthew effect*, the *trap of routine assessment*, and the *toxic power of data and experiments*—offer a set of grand challenges that learning-at-scale tinkerers will need to tackle in order to see more dramatic improvements in school systems. (This paper is adapted from the introduction to [38].)

Author Keywords

Synthesis, MOOCs, adaptive tutors, peer learning, learning at scale

CSS Concepts

• **Human-centered computing~Human computer interaction (HCI); Education**, Interactive learning

The late 2000s and 2010s were marked by a series of bold, unfulfilled predictions for learning at scale. In 2008, Harvard Business School professor Clayton Christensen, with colleagues Curtis Johnson and Michael Horn, wrote a book called *Disrupting Class* about online learning and the future of K–12 schools. They predicted that in ten years—by 2019—half of all middle and high school courses would be replaced by adaptive, self-paced online courses, “the cost will be one-third of today’s costs, and the courses will be much better.” [9] Salman Khan founded Khan Academy in

2007 to develop these kinds of online materials [19]. His vision captured the attention of popular media, and *Wired*, *Time*, and *Forbes* featured Khan on their covers with stories like “One Man, One Computer, 10 Million Students: How Khan Academy is Reinventing Education” [29, 34]. Khan’s own book was titled *One World Schoolhouse: Education Reimagined*, and as a demonstration school for these transformational ideas, he founded the Khan Lab Academy [20].

Some reformers went further, arguing that in an internet-connected world, schools weren’t even necessary. In a 2013 TED talk and a 2016 ACM Learning@Scale conference keynote, Sugata Mitra argued that with basic internet access and limited adult intervention, young people could teach themselves any topic, asserting that his two key initiatives—Hole in the Wall, and School in the Cloud—could transform learning around the world [31, 32]. In his 2016 Learning@Scale abstract, Mitra declared without reservation, “*Thirteen years of experiments in children’s education takes us through a series of startling results - children can self organise their own learning, they can achieve educational objectives on their own, can read by themselves. Finally, the most startling of them all: Groups of children with access to the Internet can learn anything by themselves.*”

In higher education, massive open online courses (MOOCs) were ground zero for the rhetoric of disruption, transformation, and renewal. In 2011, Peter Norvig and Sebastian Thrun—researchers with affiliations with Google and Stanford—offered an online course called Introduction to AI, with short online videos and interspersed practice problems inspired by Khan Academy [18, 29]. When over 160,000 learners signed up to participate in the course, elite higher education institutions embraced online learning with staggering speed. Stanford spun out the for-profit MOOC providers Coursera and Udacity, and Harvard and MIT created a non-profit MOOC provider, edX. Millions signed up for the first new offerings from these institutions, the *New York Times* declared 2012 the “Year of the MOOC,” and MOOC providers promised a radical reconfiguration of higher education [37]. In August 2013, Sebastian Thrun, Udacity’s founder, declared, “The thing I’m insanely proud of right now is I think we’ve found the magic formula” for offering powerful learning experiences at low cost and global scale [8].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

L@S '20, August 12–14, 2020, Virtual Event, USA

© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-7951-9/20/08...\$15.00

<https://doi.org/10.1145/3386527.3405929>

1.0 THE CHARISMATIC AND THE TINKERER: TWO STANCES FOR LEARNING AT SCALE

The projects of Christensen, Khan, Mitra, and Thrun exemplify what anthropologist Morgan Ames describes in *The Charisma Machine*, her history of the One Laptop Per Child initiative, as the “charismatic” approach to education technology [2]. The charismatic stance ascribes tremendous power to new technologies to reinvent education. In the 2000s and 2010s, charismatic technologists often adopted the rhetoric of “disruptive innovation” to describe how emerging technologies can offer a new value proposition that leads to the wholesale transformation of existing systems. The TED conference hosts the annual revival meeting for charismatic technologists, and reading transcripts from TED talks about educational topics is a reliable research strategy for finding unfulfilled predictions about education technology over the last decade.

One opposing position to the charismatic stance is skepticism, and education technology has a rich tradition of critique that designers and educators should take seriously (see, for instance, [40] and [51]). But as an alternative to the charismatic, Ames offers the *tinkerer*, a term drawn from David Tyack and Larry Cuban’s history of K–12 schooling in the United States, *Tinkering toward Utopia* [47]. Tinkerers see schools and universities as complex systems that can be improved, but they believe improvements come from many years of incremental changes to existing institutions rather than from wholesale renewal. Tinkerers harbor an optimism that technology can be used to improve teaching and learning, but they embrace research and critique as a crucial check against utopian thinking. Tinkerers study past efforts at educational reform to avoid replicating past mistakes. Charismatic technologists orchestrate boom-and-bust hype cycles; they cajole local systems into making major changes and then move on when transformation prove elusive. Tinkerers persist much longer with their designs, their partners, and their communities.

Millions of people have watched Sal Khan’s two TED talks, but I suspect that far fewer have read his 2019 interview with *District Administrator* magazine, a little trade journal for school superintendents and central office staff, in which he told the interviewer, “Now that I run a school, I see that some of the stuff is not as easy to accomplish compared to how it sounds theoretically” [7]. In 2019, Khan’s recommendations focused less on disruptive transformations of math education and more on using Khan Academy as modest supplement to traditional classroom instruction:

More recently, we’re seeing that if students put 30 minutes to an hour per week—or one class period per week—toward software-based, self-paced learning, schools will see a 20 to 30 percent greater-than-expected gain on state assessments. That’s exciting because that’s a dosage that’s very doable in mainstream classrooms. We tell schools to give

students 30 to 60 minutes of Khan Academy per week, with teachers doing traditional curriculum four days per week. You’re going to see a pretty dramatic improvement. You’ll get the best of both worlds. [7]

When Khan’s soaring vision met the complex reality of schools, disruption and transformation gave way to accommodation.

The blended model that Khan espoused in *District Administration* is actually at least 25 years old. In 1997, Ken Koedinger and colleagues published “Intelligent Tutoring Goes to the Big City,” a report in the *International Journal of Artificial Intelligence in Education* that described the use of adaptive, self-paced algebra tutoring software in the Pittsburgh public schools. Students learned in traditional settings for most of their class time and then spent about one day a week using math tutoring software, which led to improvements on math tests similar to what Khan found decades later [22]. Much of what Khan Academy discovered in 2019 about computer-assisted math instruction—after more than \$100 million in philanthropic investment—could have been found in academic papers published in the 1990s.

In Khan’s *District Administration* interview, I see a welcome turn from the charismatic stance to the tinkering stance, from an effort to use technology to disrupt schools to the recognition that schools are complex, conservative institutions, and the efforts of technologists, designers, and education reformers are better spent tinkering—targeting specific improvements in specific areas—than pursuing magic formulas or all-knowing robot tutors in the sky.

One of the tinkerer’s most useful resources is the historical and scholarly literature on schooling and technology in schools (see [11]). Very few emerging large-scale learning technologies are wholly new. Education technologists have been building systems for using computers to teach students since the days when mainframe computers took up entire rooms [3, 12]. Digital learning at scale is a 60-year-old endeavor; distance education goes back at least a century. At their best and most innovative, most new technologies offer an incremental improvement over past developments. Tinkerers embrace this incremental improvement as the most feasible pathway toward meaningful improvement and change in learning institutions.

Looking back at the past is made more challenging by the insistence of technology advocates and enthusiasts that their new inventions are wholly new, disruptive, and transformative. To help synthesize this long history of large-scale learning technologies, I offer two contributions. First, I organize learning-at-scale technologies into three genres based primarily on the agent that determines the next action in a sequence of learner actions. I call these *instructor-guided*, *algorithm-guided*, and *peer-guided* genres of learning at scale. If you can figure out how a new large-scale learning technology fits into one of these three

Three Genres & Four Dilemmas in Learning at Scale

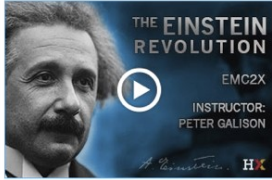
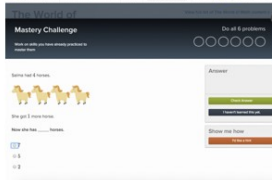
TEACHER ALGO PEER	<u>Pedagogy</u>		<u>Technology</u>	<u>As-Yet Intractable Dilemmas</u>
	 xMOOC	 Thorndike	LMS+ Autograders	Curse of the Familiar
	 Intelligent Tutors	 Dewey	IRT+ Autograders	EdTech Mathew Effect Trap of Routine Assessment
	 cMOOC, Scratch		Web+ Aggregators	Toxic Power of Data and Experiments

Figure 1: Three Genres and Four Dilemmas of Learning at Scale

genres, you can do two useful things. First, based on the prior performance of similar approaches, you can make predictions about what outcomes will emerge as new technologies are integrated into complex systems of schooling. Second, understanding what is old and recycled can throw into relief what is genuinely innovative in a new product or approach. Identifying the modest innovations in new technologies can help predict how a new offering might offer some incremental improvement over past efforts.

Second, I identify four kinds of obstacles that large-scale learning systems have encountered repeatedly over recent decades, which I call “as-yet intractable dilemmas”: *the curse of the familiar*, *the edtech Matthew effect*, *the trap of routine assessment*, and *the toxic power of data and experiments*. These patterns help explain why predictions for transformative change have fallen flat, and examining these dilemmas is critical for looking forward. These four dilemmas are a set of grand challenges that designers, researchers, and educators must overcome in order for education technology to make learning faster, cheaper, more enjoyable, more effective, and more accessible for people around the world. In what follows, I provide an introduction to these three genres and four dilemmas as a framework for organizing the history of learning at scale in a way that could inspire in the 2020s more productive approaches for improvement and innovation than we saw in the 2010s.

2.0 THREE GENRES OF LEARNING AT SCALE: INSTRUCTOR-GUIDED, ALGORITHM-GUIDED, AND PEER-GUIDED LEARNING AT SCALE

Large-scale learning environments can be classified into three genres, based on the answer to the question, Who creates the activity sequence for learners? These sequences can be created by instructors (as in the case of MOOCs), by algorithms (as in the case of adaptive tutoring software), or by peers (as in the case of distributed learning networks). Each of these genres of instructor-guided, algorithm-guided, and peer-guided large-scale learning technologies has a history, a research literature, and a track record of success and failure in formal educational institutions. Each genre also uses a common set of core technologies, and they reenact pedagogical debates that have deep roots in the history of education (Figure 1 summarizes the three genres).

The MOOCs created by elite universities are examples of instructor-driven learning experiences. Instructors design or select lectures, readings, and activities that form a knowledge base for student learning [5]. Learners are assessed by tools and systems designed by instructors that can range from simple multiple-choice questions to complex systems for evaluating computer programming assignments [27]. The learning experiences in the course are arranged in a particular order—from the Shang dynasty to Mao’s People’s Republic, say, or from “Hello World” to recursive algorithms—that are selected by the instructor. A

student may be free to traverse this material in her own way—and she might help a peer along the path—but most students generally proceed along the main path laid out by instructors.

Adaptive, large-scale learning environments are those where each item in a learning sequence is selected by an algorithm or other system on the basis of student performance in previous parts of a learning sequence. These kinds of learning experiences are often called *adaptive tutors* [24] or *computer-assisted instruction*. Khan Academy offers a useful example. While Khan Academy is best known for Khan’s video lectures, when Khan Academy is used in schools, students spend 85% of their time doing practice problems [35]. These problems will be familiar to anyone who has ever completed a worksheet in math class; they pose a question, and students must provide a correct answer by inputting an equation, selecting a point on a Cartesian plane, ordering a series of numbers in a line, or selecting from a list of multiple-choice options. The problems are organized into topics, such as dividing fractions or solving quadratic equations.

Unlike a paper worksheet, however, with an adaptive tutor, the order of problems that a student encounters depends upon her performance on each problem. Within a class of problems—such as multiplying fractions—some problems are easier (multiplying by $\frac{1}{2}$) and some are harder (multiplying by $\frac{3}{4}$). Students are given an initial problem, and if the student gets a problem right, an algorithm assigns a more difficult problem. If she gets it wrong, the system assigns an easier problem, perhaps with some form of remediation, like a hint or a link to an explanation. These systems are often called adaptive because they can increase or decrease in difficulty and provide specific remediation based on the performance of the student. In nearly all MOOCs from edX or Coursera, every student receives the same number of problems and assignments, which are presented in the same order. Students using Khan Academy and other adaptive tutors are offered a set of assignments that are dynamically adjusted for the individual student.

In peer-driven learning environments—like those proposed by Sugata Mitra in the School in the Cloud—participants can offer instruction, examples, comments, and feedback, and users can follow each other and form subgroups and networks. Mitra argued that if learners were organized into small groups with access to the learning resources of the internet and some minimal on-demand mentoring and coaching (he proposed using a network of British pensioners in his trials), then students could learn any topic of any complexity [32, 33]. The original connectivist-inspired MOOCs (cMOOCs) provide another example of a peer-guided large-scale learning community. Participants created their own blogs, social media accounts, and other sites on the open web, where they responded to course prompts and to each other. Instructors used the course home page and other technologies to aggregated copies of these

diverse contributions into one central location. But at their most successful, peer interactions were the driving force of cMOOCs [26, 31].

The most prominent peer-driven learning environment in K–12 schools is the community organized around the Scratch programming language, developed by the Lifelong Kindergarten Lab at MIT [39]. Scratch is a block-based programming language in which the young and young-at-heart can learn to program by dragging “blocks” with executable code instructions into place with other blocks, rather than by writing programming syntax with specifications for spacing, semicolons, variable names, and so forth. By default, all Scratch programs exist as projects, all projects are publicly viewable and openly licensed, and all projects can be forked and remixed as new projects; sharing and community are thus integral parts of the experience of using the Scratch programming language. In these communities, there are designers and leaders; Mitch Resnick, Natalie Rusk, and many others in the Lifelong Kindergarten Lab create the environment for Scratchers to work and learn, highlight projects on the Scratch website and social media, and cultivate community. This community then creates a wide array of projects, tutorials, guides, and other subcommunities, and learners in the Scratch community then choose for themselves how they navigate this web of opportunities for practice and learning.

I define these three categories—instructor-driven, algorithm-driven, and peer-driven—as genres of learning at scale. Hybridity between categories and uniqueness beyond these three categories certainly exist, but much of the existing infrastructure of large-scale online learning fits in an instructor-driven, algorithm-driven, or peer-driven bucket. Within each of these genres, large-scale learning tools use similar technology architectures, and they embrace similar pedagogical approaches. By identifying which of these three slots a new technology or service offering fits into, education stakeholders can make good bets about how it’s built, how it will likely be used to support learning, and what historical analogs might exist that can help predict its impact.

2.1 Core Technologies of Learning at Scale Genres

Within the three genres of learning at scale, common patterns in technology architecture emerge. Instructor-driven learning environments typically build off learning management systems and autograders; algorithm-driven environments build off autograders and adaptive algorithms; and peer-driven systems use the open web with navigational aggregators like search terms and hashtags. Many emerging large-scale learning technologies will include variants and improvements on these technologies, but usually, the roots of these technologies are decades deep. Situating new technologies within these patterns helps predict how they will perform and how they will be integrated into complex systems.

2.1.1 Instructor-Guided Learning at Scale Technologies

Instructor-paced systems require a mechanism that presents content to learners in an instructor-defined order while simultaneously tracking individual participant progress through that content. Learners must be able to start and stop at their own pace and then pick up right where they left off. The core platform enabling most instructor-paced systems is the learning management system (LMS). The phrase *management system* has been used to describe computer-based instructional platforms since the 1960s, years before the advent of personal computers [4, 10]. Since then, LMSs have become ubiquitous in higher education and many K–12 schools as storage and dissemination systems in residential learning environments and as a complete delivery platform for fully online learning. Instructors place readings, course notes, lecture videos, syllabi, and assignment prompts online, usually following linear structures that map onto the weekly schedule of typical residential courses. Students might be required to contribute to topical discussion forums, to submit their assignments in online dropboxes, and to take short multiple-choice quizzes or surveys directly inside the LMS platform, but these interactive features are typically secondary to the role of the LMS in content dissemination.

For an LMS to support the certification of learning at scale, it needs a second key technology: autograders. In courses with thousands of students and one or two instructors, if the course aspires to certify or credential learners, it needs some mechanism to assess performance at scale. The autograding tools in MOOC platforms include both multiple-choice quiz questions and a variety of forms of quantitative and computational input. Students can choose a multiple choice answer for some problems, enter numerical answers, algebraic equations, and chemical formulas for others, and submit short computer programs to be evaluated and tested by other computer programs.

These grading mechanisms depend principally upon pattern matching, where course designers program a bank of possible correct answers and computationally compare student responses to these banks. The earliest versions of these pattern-matching systems were programmed not computationally but mechanically. Sydney Pressey, a professor at Ohio State University, developed a “teaching machine,” patented in 1928, that presented students with a series of multiple-choice questions with four possible answers, each of which advanced to the next question only upon submission of a correct answer (entered by touching a typewriter-like key) [50]. From the earliest days of mainframe computers in the 1950s, technologists have sought to adapt computers to similar functions. The earliest forms of computer-assisted instruction maintained banks of words, phrases, or numbers that represented correct responses to a question (“5,” “five” “Five,” “fiv”), which were computationally compared to student responses. Even complex contemporary efforts like autograding natural-

language essays boil down to mathematically intensive pattern matching rather than sense making.

2.1.2 Algorithm-Guided Learning at Scale Technologies

The genre of algorithm-driven learning environments also requires autograders, but these technologies do not use a learning management system that sequences content in a linear order. Rather, autograders assess human performance, and then an algorithm determines whether to assign an easier problem or a harder problem, whether to certify that a learner has “mastered” a topic and is therefore ready for a new sequence of problems, or whether some other instructional step is appropriate. For these processes to work, the system must be able to maintain models of student proficiency and question difficulty, which it can then adaptively match. Fortunately, this is another problem in education technology that designers and researchers have worked on for decades. Researchers in the 1950s and 60s at Educational Testing Services and elsewhere developed a statistical toolkit—item response theory—for creating a model of question difficulty [49].

Item response theory emerged to solve a basic problem in test design. When testing large numbers of students, consumers of testing data (admissions offices, employers, policymakers) would like to be able to compare two students tested on the same subject area. Testing companies, however, would prefer not to test all students on the exact same material, since using identical test items and formats with different students in different places and times opens the door to cheating and malfeasance. Giving students different tests whose results can be fairly compared requires the ability to quantify the difficulty of each individual question and then compare each against all others.

A model of item difficulty can be used to fairly compare test outcomes, and it can also be used to provide students with progressively more or less difficult items from a bank, based on their performance. Algorithmically driven learning environments combine autograders with these models of item difficulty to create learning environments that are responsive to student performance. Just as learning management systems and autograders have roots that go back decades, so do the algorithms that model item difficulty and student performance.

2.1.3 Peer-Guided Learning at Scale Technologies

Peer-driven learning environments are powered by technologies that enable networked learning. In the early 1970s, a compelling critic of formal schooling, Ivan Illich, imagined the kinds of technologies required for peer-driven online communities. He wrote:

A student who has picked up Greek before her vacation would like to discuss Greek Cretan politics when she returns. A Mexican in New York wants to find other readers of the paper *Siempre*—or of “Los Agachados,” the most popular comic book. Somebody

else wants to meet peers who, like himself, would like to increase their interest in the work of James Baldwin or of Bolivar.

The operation of a peer-matching network would be simple. The user would identify himself by name and address and describe the activity for which he sought a peer. A computer would send him back the names and addresses of all those who had inserted the same description. It is amazing that such a simple utility has never been used on a broad scale for publicly valued activity.

In its most rudimentary form, communication between client and computer could be established by return mail. In big cities typewriter terminals could provide instantaneous responses. The only way to retrieve a name and address from the computer would be to list an activity for which a peer was sought. People using the system would become known only to their potential peers [16].

Though the technologies of today are not quite what Illich imagined, his vision of “learning webs” offers an uncanny description of the essential features of networked learning in the internet era. Illich’s fundamental metaphor, the bazaar, constrained his imagination to an expanding series of small—ideally dyadic—tutorial interactions. But most peer-driven learning environments occur either on the open web or on web-like open platforms, which allow all kinds of interactions between individual learners, peers, and experts of varying levels of proficiency and formal designation. Search engines and keywords—rather than punch cards and mainframes—function as aggregators and provide the essential connections between tutorial makers and learners.

2.1.4 Classifying Learning-at-Scale Genres by Technology

Understanding how genres of large-scale learning technologies build on common technological infrastructure provides three useful insights. First, recognizing these familiar technologies helps classify emerging technologies into the three genres. Second, familiarity with the history of these technologies evaluates and critiques the claims of novelty that are so often deployed by charismatic technologists. Jose Ferreira, founder of the edtech company Knewton, claimed that his adaptive tutoring tools were “like a robot tutor in the sky that can semi-read your mind and figure out what your strengths and weaknesses are, down to the percentile” [52]. As Ferreira was describing a system of unprecedented scale and complexity, Knewton engineers were simultaneously publishing blog posts with titles like “Understanding Student Performance with Item Response Theory.” In one post, engineers declared, “At Knewton, we’ve found IRT models to be extremely helpful when trying to understand our students’ abilities by examining their test performance” [25]. Lift up the hood of the magical robot tutor, and underneath was a 40-year-old

technology powering the whole operation. Understanding this continuity helps critics and tinkerers ask the question, What’s really new here, and can the specific elements of novelty actually produce the dramatic results claimed by evangelists?

While researchers and developers have been working on computer-based learning technologies for over half a century, educators have been developing pedagogical ideas for millennia. If the technological underpinnings of most new edtech tools are starting to look a little old, the pedagogies underneath them have roots that go back millennia.

2.2 New Technologies, Old Pedagogies

In the essay “On Listening,” Plutarch wrote that “education is not the filling of a pail, but the lighting of a fire” [43]. Of course, theories of learning and instruction can be infinitely more complex than this simple dichotomy, but these two perspectives—bucket filling and fire lighting—appear in various guises throughout education history, perhaps nowhere more clearly delineated than in the contrasting theories of Edward Thorndike and John Dewey. Ellen Lagemann, the former dean of the Harvard Graduate School of Education, summarized the last century of educational history this way: “I often argued to my students, only in part to be perverse, that one cannot understand the history of education in the United States during the twentieth century, unless one realizes that Edward L. Thorndike won and John Dewey lost” [28].

John Dewey, perhaps the foremost pedagogical progressive in the United States, was a philosopher of the pragmatist school. He famously argued that “education . . . is a process of living and not a preparation for future living,” and he advocated for an approach to education that emphasized apprenticeship, interdisciplinary learning, and connections to the world beyond schools [13]. His ideas spawned pedagogies and pedagogical philosophies whose variants might reasonably be categorized as expressions of *social constructivism*—the idea that individuals construct new understandings from prior understandings in the context of social groups.

Edward Thorndike is less well-known than Dewey, but his approach to education will be easily recognized. An early developer and advocate of standardized tests and intelligence testing, Thorndike believed that education could be organized as a science—in the positivist traditions that shaped sociology, political science, economics, and other social sciences—and that learning could be precisely measured [46]. With these measures of learning, educational techniques could be evaluated to identify those that were most and least effective at promoting learning gains on standardized measures, and best practices could be standardized and scientifically evaluated. The intellectual descendants of Thorndike’s thinking can be generally grouped as expressions of *instructionism*—learning with an emphasis on direct instruction from experts to novices.

In the decades since Thorndike and Dewey, educational theorists have offered updated approaches to both philosophies (constructionism [36] and connectivism [41] for the social constructivists, and cognitive load theory [44] for the instructionists) and have developed new approaches based on syntheses of these ideas (such as the knowledge-learning-instruction framework [23]). For all of these innovations, Lagemann's point still holds: Thorndike and Dewey have come to define two poles of an educational debate in the American context. If Plutarch were to declare to both of them that education is not the filling of a pail but the kindling of a flame, Dewey would nod, and Thorndike would scoff while measuring the flow rate into the pail.

Rather than resolving this pedagogical debate, learning at scale reenacts it.

Both philosophies—the bucket-filling and the fire-kindling—are at play in today's large-scale learning environments. Khan Academy's adaptive technologies depend upon quantifying the state of student knowledge in a domain (like multiplying fractions) and the difficulty of a given problem, such that the student can be dosed with the optimal level of difficulty for each problem. University MOOCs are motivated by the utilitarian challenge of providing the maximum number and diversity of learners across the world with access to content produced by the world's most esteemed thinkers in a given field. Learners receive knowledge primarily through lectures, and they recite such knowledge primarily through multiple-choice and quantitative tests. The self-paced and instructor-paced genres of learning at scale tend to enact a Thorndike-inspired “banking” model of education.

By contrast, designers of peer-driven learning environments are eager to create learning environments that mirror as closely as possible the knowledge-building practices used by professionals in fields and disciplines. The act of building a network of resources and fellow learners is more important than any specific (and certainly any measurable) act of learning. The goal is not to optimize the trajectory toward a defined end, but to mirror learning practices of the real world. Though Dewey and Thorndike would likely be befuddled by Scratchers building new programs on flat tablet computers, they would recognize the apprenticeship model of novices looking over the shoulders of journeymen, watching them work and trying new things themselves.

2.3 Outcomes

Scholarly literature on each of the three genres offers valuable insights about what might be predicted from new entrants. The most well-developed literature, in terms of meta-analysis, is for the adaptive tutors in algorithm-guided learning at scale. These tools have been used primarily in subjects where autograders can evaluate well-defined right and wrong answers (in the K–12 curriculum, these typically reside in math and the early parts of the reading curriculum). Meta analyses, especially those by economists

and others outside the education technology field, have generally found null average results for the use of reading tutors and modest positive results for math tutors [14, 42]. Of course, it always remains possible that an emerging technology could lead to dramatic changes. But when read in the light of the scholarly literature, Khan Academy's predictions for transformation in the early 2010s seem like long odds.

A long history of distance-education research shows that instructor-guided self-paced online learning experiences can be used by some students to successfully pursue higher education, but persistence and completion are serious problems [6]. MOOC researchers over the last ten years have found similar results, with the important wrinkle that self-regulated learning strategies appear to be essential for persistence and success, and these self-regulated learning skills are often correlated with measures of educational attainment and socioeconomic status [1, 21]. Instructor-guided self-paced online learning appears to work best for already-educated, already-affluent learners. Researchers summarizing the literature on online learning in K–12 have come to similar conclusions: high achieving learners are most likely to benefit from online options [15].

Researchers in peer-guided learning at scale have been less likely than researchers studying the other two genres to use quantitative research, preferring instead to produce rich qualitative descriptions of successful learners. Since learners bring their own learning goals to peer-driven learning environments, standardized measures of average achievement have often not been the methodology of choice for those studying peer-learning communities. Two key themes that emerge from the qualitative literature on learners in peer-driven communities is that successful learners are quite extraordinary, demonstrating a remarkable capacity to synthesize new insights, skills, and knowledge from peers online. The most sophisticated Scratchers, for example, produce animations, games, and other programs that showcase a sophisticated understanding of computer science and computational thinking [39]. Many learners, however, struggle to find their footing in peer-driven learning environments. People may get bored going through xMOOCs and cognitive tutors and may eventually quit for lack of sustained interest, but xMOOC students don't tend to become so confused that they don't know how to participate. By contrast, many cMOOC participants find these networked learning environments overwhelming, and students' getting stuck—not knowing what to do next to advance their learning—is a common challenge in classrooms adopting Scratch. One of the signature design challenges of peer-guided learning environments is to figure out how to make them more accessible to novices without turning them into instructor-led learning environments [17, 27, 33].

Nowhere in this overview of large-scale learning technologies is there an example of a technology that

proved disruptive to existing systems or that provided transformative benefits to an entire system of learners. Rather, the literature shows that the best designed learning technologies serve particular people in specific niches of existing education systems; they work, they have value, but they do not lead to the wholesale transformation of systems. The limits of their impact are gated by a set of dilemmas that cut across all three genres of learning at scale.

3.0 AS-YET INTRACTABLE DILEMMAS IN LEARNING AT SCALE

There are many differences between the progressive, open-ended, peer-driven Scratch community and the linear, instructor-driven catalog of MOOCs. Yet both learning systems are bound by similar struggles to minimize inequality and to balance innovation in design and purpose with the limits of learners' tolerance for novelty. For all the differences between these three genres of learning at scale, nearly all large-scale learning technologies face a common set of challenges in reshaping educational systems or even just serving existing systems. Typically, the three genres of learning at scale are studied by educational researchers from different communities: researchers studying algorithm-driven systems attend the Artificial Intelligence in Education or Educational Data Mining conferences; researchers interested in instructor-driven systems attend Learning Analytics and Knowledge or Learning with MOOCs; and designers building and researching peer-driven systems attend the Connected Learning Summit. Since some of the most pressing dilemmas in learning at scale are common across the three genres, effective strategies for managing these obstacles that are discovered within one genre might very well be effective across the field. Although I view these dilemmas as persistent, durable, and resistant to simple technological fixes, I am an optimist at heart (albeit one chastened by the history of education technology), and I believe they can be overcome by equivalently persistent designers, reformers, and practitioners.

There are four kinds of obstacles—I call them “as-yet intractable dilemmas”—that large-scale learning systems have encountered repeatedly over recent decades. These challenges are useful in looking backward; they help explain why grand predictions for transformative change fall flat. But I also hope they are useful for looking forward; these are the obstacles that designers and educators must overcome in order for education technology to reach its potential. I call these dilemmas *the curse of the familiar*, *the edtech Matthew effect*, *the trap of routine assessment*, and *the toxic power of data and experiments*.

3.1 The Curse of the Familiar

The conservatism and complexity of schools and learners elicits a set of reinforcing tendencies in edtech developers. Schools have many recognizable, durable features: schedules, desks, worksheets, bells, credit-hours, and other features that David Tyack and William Tobin call the “grammar of schooling” [48]. The curse of the familiar is

that the easiest way to get a new technology adopted is to digitize one of these existing elements. One of the most widely used education technology websites in the world—called Quizlet—allows users to create, share, and use digital flashcards. Quizlet was able to reach millions of students in the US and around the world because it created a learning experience instantly familiar to anyone who has ever studied factual content using index cards. But if we gathered educational experts from around the world to describe the most pressing challenges faced by school systems, I suspect that “limited flashcard availability” would not make the list of major concerns. Digitizing flash cards is unlikely to be an important part of solving complex problems of education achievement, opportunity, or inequality. When developers do create new technologies that offer substantially new approaches to teaching and learning—especially the peer-guided initiatives such as cMOOCs or Scratch—educators often perceive these offerings as confusing or irrelevant. As a result, truly novel approaches to technology-mediated teaching and learning have to be paired with substantial efforts to help educators and learners understand how to reshape existing systems to allow new technologies to be useful. Education technology needs education reform more than education reform needs education technology.

3.2 The Edtech Matthew Effect

The Matthew effect is a sociological theory drawn from the biblical book of Matthew: “For to him who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away.” The edtech Matthew effect is that new technologies—even those that are available freely and openly on the internet—are typically used more often and more efficaciously by already-affluent learners, so new technologies tend to exacerbate rather than ameliorate inequalities in educational opportunities. Moreover, the barriers to equity are often less about access to technology, and more about access to robust networks of human support that can make learning technology meaningful. In order to prevent learning at scale from making already yawning opportunity gaps even worse, developers, policymakers, and educators need to develop a set of design principles for digital equity—guidelines for making technology-mediated learning that doesn't simply reinforce existing inequality.

3.3 The Trap of Routine Assessment

The trap of routine assessment focuses in on one of the critical sources of unevenness in education technology. Computers excel at anything that can be reduced to a series of routines or to pattern matching exercises. From this principle, two corollaries emerge, one in assessment technologies and one in the labor market. Our computer assessment technologies are very good at assessing routine tasks with well-defined right answers. Similarly, in the labor market, mechanical robots and online bots will take over much of the routine work that we currently pay people to do. Put another way, our assessment technologies are

particularly good at assessing the kinds of human performance that we no longer need humans to perform. To the extent that we use standardized, computationally autograded tests to guide our education system, we will increasingly test and teach students in domains where human work is either unnecessary unrewarded. Technologists may believe that with advances in artificial intelligence or machine learning, these limits will be overcome in the years ahead. But there are good reasons to believe that to the extent that learning at scale depends upon automated assessment, many important domains of human learning will remain beyond the boundaries of what can be taught and assessed well at scale. The usefulness and potential of large-scale learning tools closely follows the contours of what can be meaningfully assessed by computers.

3.4 The Toxic Power of Data and Experiments

The toxic power of data and experiments recognizes that one of the signature characteristics of large-scale learning systems is that like other online systems, they can be continuously improved through experimentation and data analysis. When children across the country complete homework assignments on paper worksheets, that data is lost to trash cans and attic boxes, and the field learns little about how students learn. When children complete homework assignments in digital environments, researchers can examine that accumulated data to understand how people learn, to assess which specific problems work well and which confuse students, and to develop instructional modifications to the system for better learning outcomes. The large-scale learning systems that have proven to be most meaningful and effective for learners have been those developed in the context of rigorous efforts at research, evaluation, and continuous improvement.

But massive data collection and experimentation are also among the most disturbing and dangerous features of what Shoshana Zuboff calls “surveillance capitalism”—systemic new efforts to monetize personal data and manipulate human behavior [53]. The data collected by online learning systems, like nuclear energy, has a potent but potentially toxic power, simultaneously capable of fueling technologies that can improve educational systems and of unleashing great harm to both individuals and society. Balancing learner autonomy, privacy, and dignity with the potential for data collection and experimentation to improve learning systems is one of the thorniest dilemmas that large-scale learning developers will face in the years ahead.

Together, these dilemmas help explain why the prophecies of edtech charismatics over the last decade have failed to come to fruition. Too often, advocates for learning-at-scale solutions have promised that new technologies would reinvent educational systems, that they could sweep away the past to make room for new futures. No doubt, there is a new generation of venture capitalists and technology advocates preparing to make similar pitches about virtual

reality or artificial intelligence or some other new technology that can be framed as a disruptive threat to the status quo. But the common theme behind the four learning-at-scale dilemmas is that the challenges to effective learning are more cultural than technological. There is no tech that erases structural inequality. There is no tech that bypasses the conservatism of education stakeholders. There is no tech that will resolve the disagreements among experts about the subjective evaluation of many facets of student work. Anyone serious about improving educational systems through technology must be equally serious about improving education policy, management, and professional learning for educators.

4.0 CONCLUSION

As we enter 2020 and a new decade of work in learning at scale, our partners in schools, colleges, policymaking, and technology development will be best served by a research community that embraces a tinkerer’s mindset. Let the lessons learned from the hype cycle of the 2010s be learned once and for all: that schools and formal education systems need partners, not disrupters. The histories embodied in the three genres and four dilemmas for learning at scale are tools for helping educators, developers, and education researchers adopt the tinkerer’s stance as patient optimists, willing to put shoulder to the wheel and engage in the long, hard, and ultimately rewarding work of incrementally improving learning through technology.

REFERENCES

- [1] M. Elena Alonso-Mencia et al. 2019. Self-regulated learning in MOOCs: Lessons learned from a literature review. *Educational Review* (2019), 1-27. <https://doi.org/10.1080/00131911.2019.1566208>
- [2] Morgan Ames. 2019. *The Charisma Machine: The Life, Death and Legacy of One Laptop Per Child*. MIT Press.
- [3] R. C. Atkinson and H. A. Wilson. 1968. Computer-assisted instruction. *Science* 162, no. 3849 (1968): 73–77.
- [4] Frank B. Baker. 1971. Computer-based instructional management systems: A first look. *Review of Educational Research* 41, 1, 51-70.
- [5] Maha Bali. 2014. MOOC pedagogy: Gleaning good practice from existing MOOCs. *Journal of Online Learning and Teaching*, 10, 1, 44.
- [6] R. M. Bernard, P. C. Abrami, Y. Lou, E. Borokhovski, A. Wade, L. Wozney, ... and B. Huang. 2004. How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research* 74, 3, 379-439.
- [7] Emily Ann Brown. 2019. Sal Khan envisions a future of active, mastery-based learning. *District Administration*, January 31, 2019. <https://districtadministration.com/sal-khan-envisions-a-future-of-active-mastery-based-learning/>

- [8] David Carr. 2013. Udacity CEO says MOOC “magic formula” emerging. *InformationWeek*, August 19, 2013.
- [9] Clayton M. Christensen, Curtis W. Johnson, and Michael B. Horn. 2008. *Disrupting Class*. McGraw-Hill, 101.
- [10] William Cooley and Robert Glaser. 1969. An information and management system for individually prescribed instruction. Working paper no. 44. *AV Communication Review* 17, 4 (1969).
- [11] Larry Cuban. 1986. *Teachers and Machines*. Teachers College Press.
- [12] Brian Dear. 2017. *The Friendly Orange Glow: The Untold Story of the PLATO System and the Dawn of Cyberculture*. Pantheon.
- [13] John Dewey. 1897. My pedagogic creed. *School Journal* 54, 3 (1897), 77-80.
<http://dewey.pragmatism.org/creed.htm>
- [14] Mark Dynarski, Roberto Agodini, Sheila Heaviside, Timothy Novak, Nancy Carey, Larissa Campuzano, and Barbara Means. 2007. Effectiveness of reading and mathematics software products: Findings from the first student cohort. National Center for Education Evaluation and Regional Assistance.
<https://files.eric.ed.gov/fulltext/ED496015.pdf>
- [15] Susan Dynarski. 2018. Online courses are harming the students who need the most help. *New York Times*, Jan. 18, 2018.
<https://www.nytimes.com/2018/01/19/business/online-courses-are-harming-the-students-who-need-the-most-help.html>
- [16] Ivan Illich. 1970. *Deschooling Society*. Harper Row.
- [17] Paulina Hauduong and Karen Brennan. 2019. Helping K–12 teachers get unstuck with Scratch: The design of an online professional learning experience. *SIGCSE '19*. <https://dl.acm.org/doi/10.1145/3287324.3287479>
- [18] Fiona Hollands and Devayani Tirthali. 2014. *MOOCs: Expectations and Reality, Full Report*. Teachers College Center for Benefit-Cost Studies of Education. Retrieved from
<https://files.eric.ed.gov/fulltext/ED547237.pdf>
- [19] Salman Khan. 2011. Let’s use video to reinvent education. Video (March 9, 2011). Retrieved from
<https://www.youtube.com/watch?v=nTFEUsudhfs>
- [20] Salman Khan. 2012. *The One World Schoolhouse: Education Reimagined*. Grand Central Publishing.
- [21] René F. Kizilcec, Mar Pérez-Sanagustín, and Jorge J. Maldonado. 2017. Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers and Education* 104 (2017), 18-33.
- [22] Kenneth R. Koedinger, John R. Anderson, William H. Hadley, and Mary A. Mark. 1997. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education* 1997, 8 (1997), 30-43.
- [23] Kenneth R. Koedinger, Julie L. Booth, and David Klahr. 2013. Instructional complexity and the science to constrain it. *Science* 342, 6161, 935-937.
- [24] Kenneth R. Koedinger and Albert T. Corbett. 2006. Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge Handbook of The Learning Sciences*. Cambridge University Press, 61–77.
- [25] Knewton Staff. 2012. The mismeasure of students: Using item response theory instead of traditional grading to assess student proficiency. *Knewton Blog*, June 7, 2012.
<https://web.archive.org/web/20130217120645/http://www.knewton.com/tech/blog/2012/06/understanding-student-performance-with-item-response-theory/>
- [26] Rita Kop. 2011. The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. *International Research Review of Online and Distance Learning*. Retrieved from
<http://www.irrodl.org/index.php/irrodl/article/view/882/1689>.
- [27] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer. 2013. Peer and self-assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 6, 1-31.
- [28] Ellen Lagemann. 1989. The plural worlds of educational research. *History of Education Quarterly* 29, 2 (1989), 185-214.
- [29] Steven Leckart. 2012. The Stanford education experiment could change higher learning forever. *Wired*, March 30, 2012.
https://www.wired.com/2012/03/ff_aiclass/
- [30] Allison Littlejohn, Nina Hood, Colin Milligan, and Paige Mustain. 2016. Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *The Internet and Higher Education* 29 (2016), 40-48
- [31] Colin Milligan, Alison Littlejohn, and Annoush Margaryan. 2013. Patterns of engagement in connectivist MOOCs. *Journal of Online Learning and Teaching* 9, 2, 149-159.
- [32] Sugata Mitra. 2013. Build a school in the Cloud. Video. Retrieved from
https://www.ted.com/talks/sugata_mitra_build_a_school_in_the_cloud

- [33] Sugata Mitra. 2016. The future of learning. *Proceedings of the 2016 ACM Learning at Scale Conference*. Retrieved from <https://dl.acm.org/doi/proceedings/10.1145/2876034>
- [34] Michael Noer. 2012. One man, one computer, 10 million students: How khan academy is reinventing education. *Forbes*, November 19, 2012. <https://www.forbes.com/sites/michaelnoer/2012/11/02/one-man-one-computer-10-million-students-how-khan-academy-is-reinventing-education/#7c96110644e0>
- [35] Robert Murphy, Lawrence Gallagher, Andrew Krumm, Jessica Mislevy, and Amy Hafter. 2014. *Research on the Use of Khan Academy in Schools*. SRI Education.
- [36] Seymour Papert. 1980. *Mindstorms: Computers, Children, and Powerful Ideas*. Basic Books.
- [37] Laura Pappano. 2012. The year of the MOOC. *New York Times*, November 2, 2012.
- [38] Justin Reich. 2020. *Failure to Disrupt: Why Technology Alone Can't Transform Education*. Harvard University Press.
- [39] Mitchel Resnick et al. 2009. Scratch: Programming for all. *Communications of the ACM* 52, 11 (2009), 60-67.
- [40] Neil Selwyn. 2013. *Distrusting Educational Technology: Critical Questions for Changing Times*. Routledge.
- [41] George Siemens. 2008. Connectivism: A learning theory for the digital age. *elearnspace*. Archived at <http://www.webcitation.org/5bCzNxTAn>.
- [42] Saiying Steenberg-Hu and Harris Cooper. 2013. A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology* 105, 4 (2013), 970.
- [43] James Johnson Sweeney. 1968. *Vision and Image: A Way of Seeing*. Simon and Schuster.
- [44] John Sweller. 2011. Cognitive load theory. *Psychology of Learning and Motivation* 55, 37-76.
- [45] Clive Thompson. 2011. How Khan Academy is changing the rules of education. *Wired*, July 15, 2011. https://www.wired.com/2011/07/ff_khan/
- [46] Edward Thorndike. 1914. *The Psychology of Learning*. Teachers College.
- [47] David B. Tyack and Larry Cuban. 1995. *Tinkering toward Utopia*. Harvard University Press.
- [48] David Tyack and William Tobin. 1994. The “grammar” of schooling: Why has it been so hard to change? *American Educational Research Journal* 31, 3 (1994), 453-479.
- [49] W. J. van der Linden. 2010. Item response theory. *International Encyclopedia of Education* 4 (2010): 81-88.
- [50] Audrey Watters. 2015. The first teaching machines. *Hack Education* (blog), February 3, 2015. <http://hackededucation.com/2015/02/03/the-first-teaching-machines>
- [51] Audrey Watters. 2020. *Hack Education*. Blog. <http://hackededucation.com>
- [52] Eric Westervelt. 2015. Meet the mind reading robot tutor in the sky. *NPR*. October 13, 2015. <https://www.npr.org/sections/ed/2015/10/13/437265231/meet-the-mind-reading-robo-tutor-in-the-sky>
- [53] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.