Pearson | SRI Series on Building Efficacy in Learning Technologies

Volume 1

Understand, Implement & Evaluate

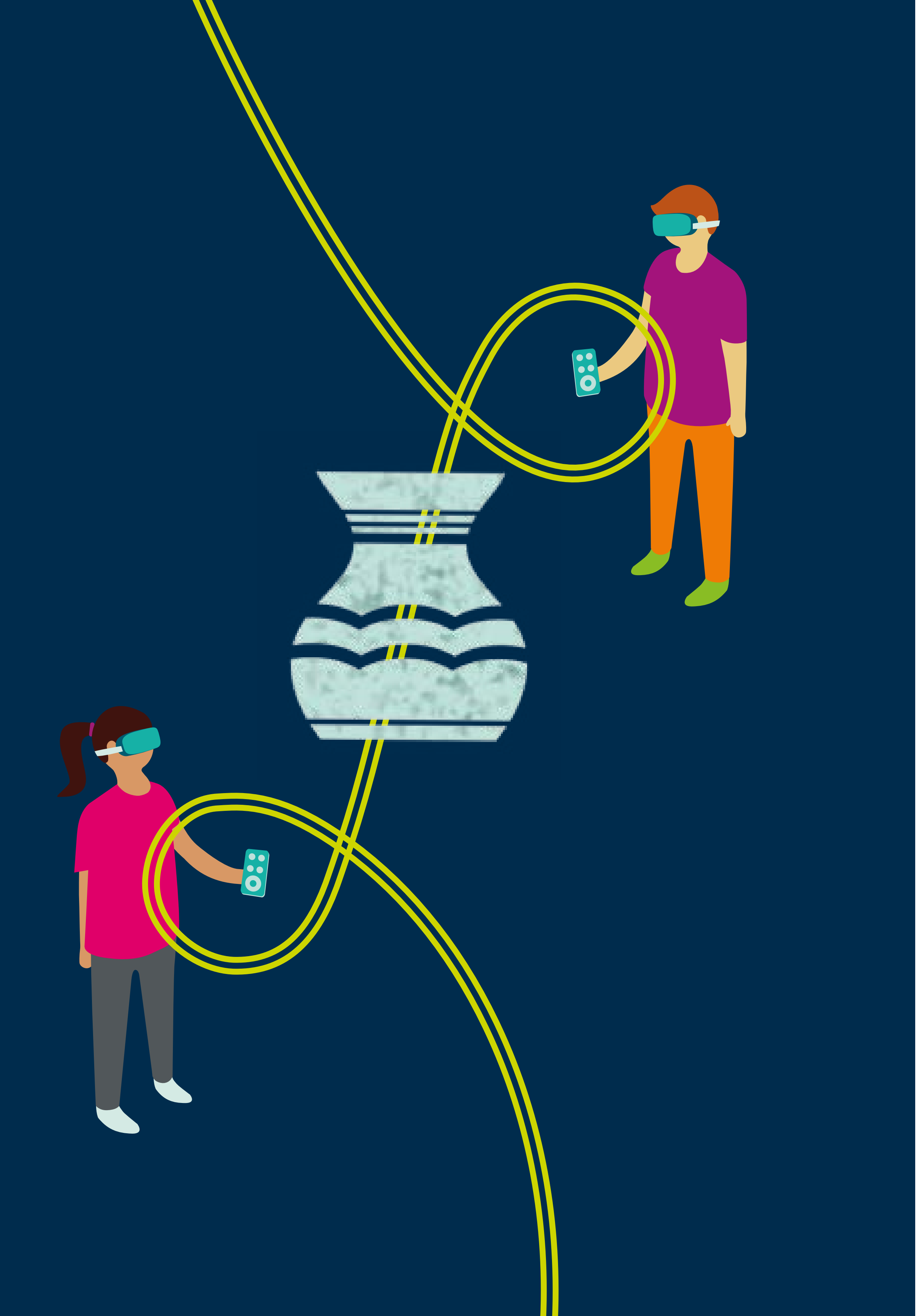
Barbara Means, Robert Murphy, Linda Shear

Open Ideas at Pearson

Sharing independent insights on the big unanswered questions in education

SRI International

Pearson



About Open Ideas at Pearson

Pearson's goal is to help people make progress in their lives through learning. This means we are always learning too.

Our Open Ideas series of publications is one of the ways in which we do this. We work with some of the best minds in education – from teachers and technologists to researchers and big thinkers – to bring independent ideas and insights to a wider audience.

How do we learn, and what keeps us motivated? What knowledge and skills do learners need as we move into the second half of the 21st century? How can smart digital technologies be used to create more personalised education? How can we build systems that offer high quality learning opportunities for everyone?

These questions are too important for the best ideas about them to stay in a lecture theatre, on a bookshelf, or alone in a classroom. They need to be found and supported, shared and debated, adopted and refined. We hope that Open Ideas helps with this, and inspires you to join the conversation.

About Pearson

Pearson is the world's learning company. We're experts in educational course ware and assessment, and provide teaching and learning services powered by technology.

We believe that learning opens up opportunities, creating fulfilling careers and better lives. So it's our mission to help people make progress in their lives through learning.

About SRI International

SRI turns bold visions into real-world products and practices.

SRI is an independent, nonprofit research center that works with clients to take the most advanced R&D from the laboratory to the marketplace. Serving government and industry, they collaborate across technical and scientific disciplines to generate real innovation and create high value for their clients.

They invent solutions that solve your most challenging problems today and look ahead to the needs of the future. For more than 70 years, they've led the discovery and design of ground-breaking products, technologies, and industries – from Siri and online banking to medical ultrasound, cancer treatments, and much more.

About the Authors

Dr. Barbara Means

Dr. Barbara Means directs the Center for Technology in Learning at SRI International, a nonprofit research organization based in Menlo Park, California. Dr. Means is an educational psychologist whose research focuses on ways in which technology can support students' learning of advanced skills and the revitalization of classrooms and schools.

A fellow of the American Educational Research Association, she is regarded as a leader in defining issues and approaches for evaluating the implementation and efficacy of technology' supported educational innovations. Currently, she leads SRI's evaluation work for the Next Generation Learning Challenge, sponsored by the Bill & Melinda Gates Foundation. Her published works include the jointly authored volumes *Learning Online, Using Technology Evaluation to Advance Student Learning, The Connected School,* and *Comparative Studies of How People Think* as well as the edited volumes *Evaluating Educational Technology, Technology and Education Reform,* and *Teaching Advanced Skills to At' Risk Students.*

Dr. Robert Murphy

Dr. Robert Murphy directs evaluation research in SRI's Center for Technology in Learning. Dr. Murphy's research focuses on the design and implementation of formative and summative evaluations of educational programs and technologies. His research combines studies of program implementation across different populations and contexts and program outcomes using experimental and quasi- experimental designs.

Currently, Dr. Murphy is co-principal investigator of a U.S. Department of Education-funded study of an online math homework platform. He is also the principal investigator of a study of online learning within programs serving low-skilled adults funded by the Joyce Foundation. Two of his recently completed projects studied how K-12 schools blending online and offline instruction to support teaching and learning, including a study of school-based implementations of the Khan Academy.

Linda Shear

Linda Shear is the director of international studies for CTL. Her research and consulting focus on systemic educational change and powerful uses of ICT (information and communications technology) to support deeper learning for all students. She recently directed research activities for the multinational Innovative Teaching and Learning (ITL) Research (http://www.sri.com/work/projects/innovative/teaching/and/learning/research), in which researchers in eight diverse countries collaborated to investigate the factors that promote innovative teaching within and across local contexts. Shear's work leverages analysis of the learning activities that teachers assign and the work that students do in response as a powerful tool for professional development. Additional areas of specialty include strategic planning and program evaluations that help clients envision and maximize the implementation effectiveness and learning outcomes of the programs they deliver.

Acknowledgements

We would like to acknowledge Pearson's support for the preparation and publication of this paper and the other two papers in the *Pearson | SRI Building Efficacy in Learning Technologies Series.* Special thanks go to Sir Michael Barber and Tim Bozik and those in their teams at Pearson, particularly Carmen Arroyo, Dan Belenky, Jeff Bergin, Kristen DiCerbo, Laurie Forcier, Mark Griffiths, Todd Hitchcock, Emily Lai, Amar Kumar, Nathan Martin, Janine Matho and David Porcaro, who provided helpful guidance as we formulated our plans for the papers as well as insightful feedback and suggestions for our earlier drafts. Thanks additionally go to Laurie Forcier and Vikki Weston, who make up the Open Ideas at Pearson team, for their work to develop this paper, and the whole series, from initial concept to finished publication.

Thanks are due also to our colleagues at the Center for Technology in Learning at SRI International. We have benefited in more ways than we can express from their expertise and commitment to research that leads to improvements in learning opportunities for everyone. The support of Denise Borders, President of SRI Education, makes our work possible.

Finally, we are indebted to the funders of our past research and the many school systems, universities, and teachers with whom we have partnered in doing research on the efficacy of innovations involving learning technologies. Without their willingness to engage with us around critical issues in education, we would not have been able to do the work that laid the experiential foundation for the perspectives presented here and in the other two papers in this series.

Readers should be aware, however, that the opinions and conclusions expressed in this paper are our own and should not be construed as necessarily representing those of Pearson or of the many organizations that have funded our research.

Suggested reference:
Means, B., Murphy, R. and Shear, L. (2017). *Pearson | SRI Series on Building Efficacy in Learning Technologies. Vol. 1 Understand, Implement & Evaluate.* London: Pearson.

ISBN: 978-0-992-42580-7

Designed by Bond & Coyne
bondandcoyne.co.uk

Contents

Foreword

At Pearson, we have made efficacy one of our company's core values. Why? First, because it's our mission to help learners make measurable progress in their lives. And second, we believe that when you use an education product you should have a clear picture of the learning outcomes you can expect, as well as an understanding of the results that others have had with that product in the past. In real terms, our commitment to efficacy means starting with learning products and services based on education research, developing a thorough understanding of how our products are used, making iterative product improvements, and generating valid and reliable claims about our products' impacts on learner outcomes through evaluative studies.

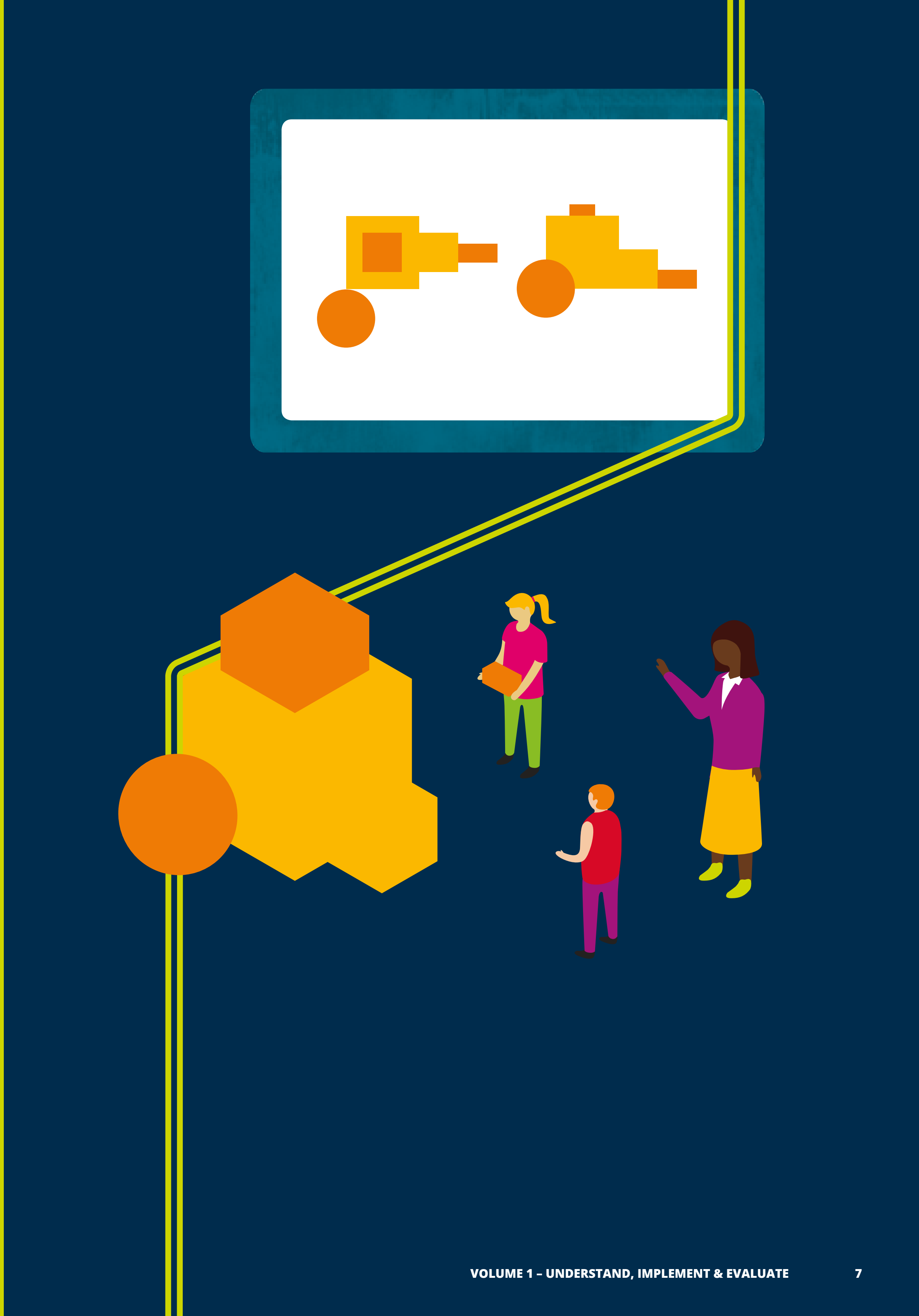
Although our commitment to efficacy spans all of our products and services, an ever-growing share of our portfolio is comprised of digital learning technologies. Understanding the efficacy of learning technologies brings its own challenges and opportunities. On the challenge side lies understanding, and accounting for, the variety of contexts in which a product might be implemented, given the profound effects that context has on impact. On the opportunity side lies the potential to use data to get "inside" learning in new and nuanced ways, which has the power to dramatically fast-forward our understanding of how people learn.

An important part of our commitment to efficacy is sharing the best of what we know from education research about incorporating learning technologies into instruction, as well as the lessons we've learned thus far. The Pearson | SRI Series on Building Efficacy in Learning Technologies represents an important element of this work. We've partnered with the experts at SRI's Center for Technology in Learning to produce this practical three-part series. In this, Volume 1, the authors helpfully contextualize the role of learning technologies within instruction, and draw out what we know thus far about the effectiveness of these tools. They then provide a step-by-step guide for identifying, planning, executing, and evaluating a learning technology in a school or school district. Subsequent volumes take on the critical issues of how data analytics can be used to improve learning technologies, and how to design learning technologies that appropriately capture actionable efficacy evidence.

We hope that you will find this series useful, and that you will join us in our drive to build efficacy into learning technologies. The demand for accessible, affordable and effective education has never been greater. Through constant measurement and improvement, we have a real opportunity to make a positive impact and reach even more learners around the globe.

Sir Michael Barber
Chief Education Advisor, Pearson

Tim Bozik
President, Global Product, Pearson



Executive Summary

In one form or another, learning technology is now found in almost every school, college, and university. Yet instructors, education administrators, teacher trainers, and policymakers struggle to find objective guidance to help them identify effective learning technology products. They would like to have a trusted resource to answer the seemingly simple question, "What works?"

This paper seeks to explain why this question is not so simple. We propose a reframing of the question into ones we can answer and that can lead to educational improvement. We will make a case for a systemic approach to learning technology implementation coupled with local iterative research to figure out what works for a particular education institution's purposes, for their learners, and in their context.

The Contentious Debate over Learning Technology Efficacy

Numerous studies demonstrate that many types of learning can be enhanced with appropriately implemented technologies. But there is also ample documentation of large-scale introductions of technology into classrooms that failed. The challenge for policymakers and educators is to figure out how to select appropriate learning technologies and implement them in ways that consistently produce positive learner outcomes.

This paper is the first in a series of three dealing with this efficacy challenge. We will treat the concept of efficacy in learning technologies in its broadest sense--encompassing issues of learning technology design and development, product selection, implementation, evaluation, refinement, and claims about impacts on learners.

The starting point for our argument is a systemic conception of learning technology implementation. Instruction should be understood as a phenomenon that emerges from the interactions among educators, students, and content. This complexity and interdependence is no less true when learning technology becomes part of the mix. Learning outcomes are shaped by the prior knowledge and actions of students and educators as well as by the quality of the technologies and other instructional materials they are given to work with.

Lessons from Research on the Effectiveness of Learning Technology

The number of controlled studies measuring the impact of learning technologies has increased markedly over the last decade. This body of research yields five important insights:

- Finding disparate results for a given learning technology product implemented in different settings is more common than not.
- It's rarely possible to disentangle the impact of a learning technology from the effectiveness of the overall instructional system in which it is embedded.
- Whether use of a digital learning product appears "effective" in a given setting depends on the purpose and goals that educators have for using technology in their classrooms.
- Even learning technology products designed for independent use are experienced differently depending on the nature and level of supports that students receive.
- When it comes to achieving learning impacts from a complex technology-enabled change in instruction, time will often be needed to iterate and learn from early experiences.

Implications for Learning Technology Implementations

These five insights have significant implications for efforts to introduce technology into instruction. The systemic nature of teaching and learning suggests that it will be necessary not only to identify an appropriate learning technology product or resource, but also to plan for the ways in which instructors and students will be changing their roles and routines to incorporate the technology and achieve coherence between technology-based activities and other aspects of instruction. This paper describes a series of mutually dependent steps in this undertaking:

Identifying promising technology tools and resources that match the goals and context for the innovation,

Planning the multiple parts of the innovation as students will experience it and the supports needed for students and instructors to be able to implement the innovation as intended,

Implementing the multiple parts of the intervention as specified in the plan,

Evaluating data to reveal how the innovation is being implemented and whether the innovation is having the desired impacts on student outcomes. Ideally, evaluation data are used to refine the technology implementation model (and sometimes the technology product itself) for future iterations of the intervention (i.e., cycling through the steps again, starting with the second step).

An Improvement Science Approach to Evaluation

Improvement science offers a set of tools and practices for systematically reflecting on processes and outcomes, trying out potential refinements, and measuring the resulting outcomes. This approach is particularly useful for efforts to achieve major educational transformations using technology. In evaluation research that incorporates improvement science, each round of process and outcome data collection feeds into analysis and reflection activities that result in refinements to the implementation model for the next iteration with a new group of students. Improvement science practitioners highlight four considerations:

- Focus on the important problem to be solved. Before engaging in improvement cycles, a group must agree upon a clearly stated long-term goal (e.g., increasing the proportion of students earning Advanced Placement credit) and a measure that can capture progress toward that goal (e.g., scores of 3 or better on the Advanced Placement examination).
- Attend to leading indicators. Because long-term outcomes take time to emerge, improvement efforts need to describe and track initial and incremental outcomes. For example, if we adopt a blended learning program, we might not see test scores improve in the first semester, but a more immediate outcome could be the changes in the teaching practices the program is supposed to catalyze.
- Success requires more than the software. As we have described, other essential ingredients of implementation include articulation of the new practices expected of educators and provision of supports for educators to learn them. One or more improvement cycles might, for example, attend to the design of the professional development that instructors receive as part of the initiative or the removal of barriers to adopting new practices.

- Consider the use of system data. Many learning technology systems provide a wealth of data about student learning paths and behaviors as well as tracking outcomes. Thoughtful incorporation of these data into improvement cycles can help instructors identify issues such as cases where students are not engaging with the learning software frequently enough to attain their learning goals.

Evaluating Impacts
Large-scale interventions affecting major portions of the core curriculum and requiring considerable investments of time and money often come with requirements for producing evidence of impact. Education decision makers for both schools and colleges should keep in mind some key points about impact studies:

- Credible impact research requires use of a comparison or control group and of a common student outcome measure. Impact studies need to have an objective learning measure common to the treatment students (experiencing the technology-supported intervention) and an equivalent comparison group.

- The best way to establish equivalence between treatment and comparison groups is to randomly assign students, instructors, or schools to treatment and comparison conditions.

- Alternatives to random-assignment experiments, which are sometimes necessary for practical reasons, can be credible if they do a good job of demonstrating the equivalence of treatment and comparison groups before the introduction of the intervention.

- Comparison of outcomes across conditions can tell you if the use of a technology-supported instructional system had an impact on learner outcomes, but not how or under what conditions.

Conclusion
This paper offers three important lessons for education leaders:

1. Efficacy is not a feature of a learning technology product per se. Products can and should be designed to leverage what we know about how people learn, but the learning technology product is always just one component of a broader learning experience. Efficacy emerges from the interactions between students, instructors, and learning activities in particular contexts. When researchers find an effect for some intervention incorporating a learning technology product, the treatment being evaluated almost always includes multiple components, such as teacher practices related to the learning technology, even if those practices were not well documented and acknowledged. Thus, the measured impact was really for the product as implemented by educators in a particular setting or settings as part of this broader constellation of practices. For this reason, the measured impact must be understood as arising from the interplay among the product's features, educator practices, and student behaviors.

2. Education leaders should take responsibility for supporting changes in instruction to get positive outcomes from the incorporation of learning technology products. When learning outcomes improve, it's almost always because core teaching and learning practices have changed. New core instructional practices only emerge when educators take responsibility for what's working and what's not, making changes to how they teach and how their students learn. Organizational supports for teacher learning and changes in practices are essential when attempting to make learning technology a core part of instruction.

3. To achieve and sustain meaningful improvements in learning outcomes, schools and colleges should measure, evaluate, and refine their instruction--repeatedly. Improved outcomes from blended learning do not arise in one swift act, but rather emerge from sustained efforts to improve and refine instructional practices over time. Educational organizations should take advantage of data for ongoing analysis of what's working well and what is not in order to refine their technology-supported instruction.

Introduction

Learning technology is at a crossroads. Research has demonstrated the potential of digital technologies to enhance learning. But we know, from experience, that technologies designed to enable learning have frequently failed to live up to that potential, falling short of the expectations of parents, teachers and education leaders. Regardless, digital technologies are affecting what and how we learn, and we need to figure out how to make them work better and more consistently. This imperative is especially strong in light of the millions of learners who either struggle when exposed to conventional instruction or who lack access to any kind of instruction on the advanced skills needed for today's global economy.

What Do We Mean By Efficacy?
A commitment to building efficacy in learning technologies means developing quality products and services based on education research, understanding how the products are used, making product improvements, and generating valid and reliable claims about the products' impacts on learner outcomes.

Given this reality, how do we build beautiful, useful educational technology that not only does "what we want it to do" in terms of functionality, but also generates evidence about whether it works for the reasons we think it should, helps uncover new insights into learning processes, exposes learning bottlenecks, and overall drives continuous improvement both in the technology and in learning and teaching?

This paper is the first in a series of three on the topic of building efficacy in learning technologies. Here in Volume 1, we first lay out our argument that any learning technology must be understood as just one part of an instructional system, not as a learning intervention unto itself. We then review what research tells us about the effectiveness of learning technologies, and make recommendations for identifying, planning, implementing, and evaluating learning technologies. Volume 2, will highlight examples of effective use of data analytics to improve learning technologies. Volume 3 will provide a roadmap for building the capability to capture actionable learning data into learning technologies from the start. Our reasons for assembling this series are twofold: one, to make the case for the importance of collecting empirical data so that we can make sound judgments about the impacts of digital learning products in teaching and learning. And two, to provide a useful toolkit to teachers, school leaders, developers, and anyone else who has a vested interest in leveraging technology for the improvement of learning and teaching, now, and in the future.

The Promise… And the Detractors…

The promise of digital technologies in education was well-captured in the seminal review of learning science research, *How People Learn*, in which the authors described education technologies that:

- Engaged real-world problems as a context for learning;
- Scaffolded portions of complex tasks and tools such as simulations and visualizations to support deeper learning;
- Provided opportunities for feedback, reflection and revision;
- Supported communications infrastructures for local and global communities of learners; and
- Expanded opportunities for educator learning.[1]

Now, nearly 20 years later, technology is far more powerful, accessible and ubiquitous in our society. The emergence of the social web and mobile computing as well as much more powerful immersive environments including multi-player online games and virtual worlds offer new promise for putting into practice conditions that research has shown enhance learning, including:

- Harnessing social aspects of learning, including collaborative learning;
- Tailoring learning content to individuals' prior knowledge, proficiency levels, and interests;
- Stimulating deeper learning that leads to retention and application in new contexts; and
- Empowering learners as producers and creators.

Moreover, people have shown that they are interested in using technology to learn new things in a new way.

Hundreds of thousands have signed up for the most popular massive open online courses (MOOCs).[2] Teachers access a wealth of online materials for lessons or professional development. Khan Academy is a global "go-to" math resource for students. Estimates of worldwide investment in educational technology topped $7 billion in 2016.[3] One initiative in this space the technology-driven "AltSchool" network of micro-schools, has received hundreds of millions of dollars in philanthropic and venture capital.

At the same time, headlines have been made by failed adoptions of large technology hardware purchases or continuing connectivity issues.[4] Meanwhile, the recent OECD report, Students, *Computers and Learning: Making the Connection*, set off a firestorm of anti-technology headlines when their data showed that increased student computer time was not associated with any improvement in PISA scores.[5]
(For more information, see box below.)

Faced with a dizzying array of new technologies, digital learning content, and instructional approaches, instructors, education administrators, teacher trainers, and policymakers seek objective guidance in identifying effective products. They would like to have a trusted resource to answer the seemingly simple question, "What works?"

Students, Computers and Learning

The OECD analysis of PISA score trends in reading, mathematics, and science found no improvement for countries that had invested heavily in technology for their schools. However, more fine-grained analyses in the same report showed that teachers using more student-centered practices, such as focusing on formulating and solving real-world problems, also used computers with their students to a greater extent. As a whole, the report suggests that associations with achievement vary depending on the way that technology is used and that schools investing in hardware need to invest also in teacher training and support.

This paper seeks to explain why this question is not so simple and proposes a reframing of the question into one we can answer and that can lead to educational improvement. We will make a case for a systemic approach to learning technology implementation coupled with local iterative research to figure out what works for a particular education institution's purposes, for their learners, and in their context.

What We Know So Far

Researchers seek to separate fact from fiction, identifying what can be claimed by technology, what is needed to evaluate those claims, and what approach is needed to create an environment where technology supports the improvement of learning outcomes.

In order for educators to be able to use technology to deliver education improvement, we need to move away from the seemingly simple question, "What works?" and towards a more systemic approach of learning technology implementation and evaluation.

Although learning technologies can certainly be of higher or lower quality, they can in no case be thought of as self-contained silver bullets ready to remedy educational ills and raise student achievement. As Fullan and Donnelly argued in *Alive in the Swamp: Assessing Digital Innovations in Education,* and Luckin and colleagues elaborated in *Intelligence Unleashed: An Argument for AI in Education,* any digital innovation in education must be understood, implemented, and evaluated as part of a larger instructional ecosystem that includes pedagogy and system change efforts involving teachers, learners, and parents.[6]

The components of that system, and the interactions among them, will shape what instructors and students do with technology and thereby the student learning outcomes—or lack thereof—that follow from the blending of digital and instructor-led instruction. This systems view of instruction has profound implications for research on the impacts of learning technology. It means that local, iterative research is needed to figure out what works for a particular education institution's purposes, for their learners, and in their context. In our experience, it is important to consider questions about the social and instructional context within which the learning technology is being used, such as:

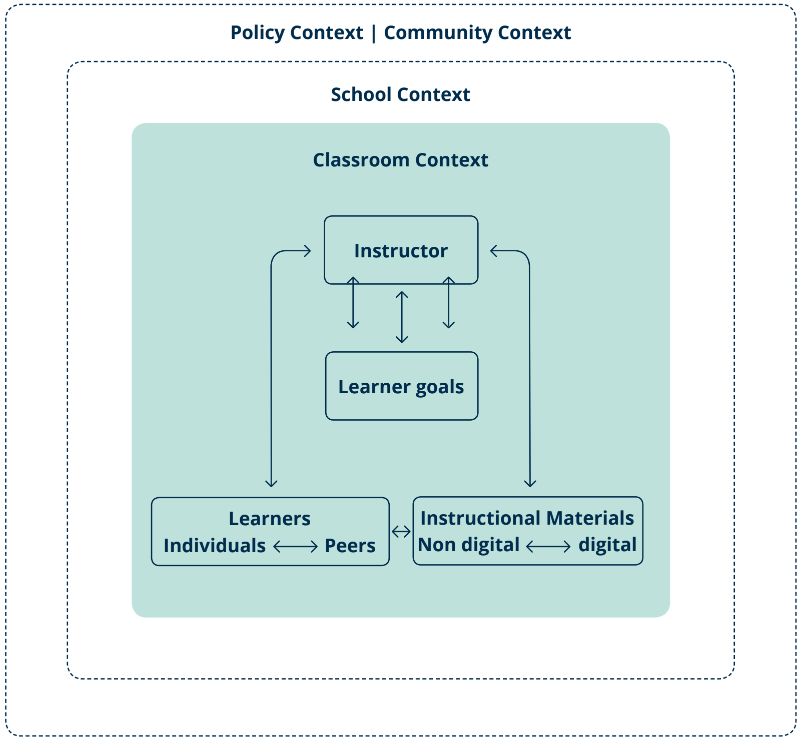
- How much time is spent in instruction on and off the technology?
- What is the content of other educator-led instruction on the same topics and its alignment with the digital learning activities?
- How do instructors introduce the technology to their students and support students as they work with it?
- What is the students' interpretation of the technology's content?
- What is their perspective on what they can gain from working with it?

Understanding Instruction

Instruction should be understood as an activity dependent on the inter-relationships among educators, students, and content.[7] The content--the material to be learned — is typically embodied in instructional materials including, but not limited to, digital learning resources.

By implication, and importantly, a learning technology product by itself is not an instructional intervention — learning outcomes are shaped by the prior knowledge and actions of students and educators as well as by the quality of the technologies and other instructional materials they are given to work with, as illustrated in *Figure 1*. From this perspective, it makes little sense to talk about measuring the learning impacts of a technology out of context. What we can measure is the impact of a broader instructional intervention that includes not only the technology but also whatever changes to instructor practices and student experiences accompany the introduction of that technology.

Fig. 1 The components of an instructional intervention.



With this paper, we first set the record straight, looking at the implications of past research on the concept of technology effectiveness. Specifically, we find that the research shows that, although learning technologies vary in quality and the amount of evidence for their use with interventions that enhanced student outcomes, *there's no such thing as a stand alone "100% effective" learning technology product or product type.*

When researchers find a positive effect of using a learning technology product, what they are really seeing is the effect of the combination of the multiple components of instruction on a particular outcome. This effect comes from a blend of contributions including the product's capabilities, the educators' practices, and the students' activities with and without the technology. This is not to say that education research findings can never be generalized. But it is to say that a documented impact must be attributed to the constellation of resources and practices present in the intervention. The nature of the learners, teachers, and measured outcomes in the study need to be considered when drawing implications for the likely effectiveness of a technology-supported intervention in other settings.

Next we will look at what we have learned about what works, showing that *getting consistently positive impacts from learning technology requires attending to the multiple aspects of the instructional system.*

When learning outcomes improve substantively, it's because the core of teaching and learning have changed. New core instructional practices emerge at scale only when education leaders and instructors take joint responsibility for identifying their desired outcomes and implementing a change in practices that is comprehensive enough to get better outcomes.

What does this look like in practice? This kind of deep change will almost always require articulating new roles for the various actors who will put the innovation into place and support it, addressing issues of teacher learning, and aligning new content with student and instructor practices, as well as making sure these new roles and practices are not undermined by continuing more familiar instructional activities.

Finally, *to achieve and sustain meaningful improvements in learning outcomes, education systems should measure, evaluate, and refine instruction, and the role of technology within that system, repeatedly.*

Outcomes from complex educational innovations do not arise in one swift act. Instead they emerge from routinized, sustained efforts to improve and refine instructional practices, as well as the support systems for executing those practices. Education systems can take advantage of learning system data as part of their efforts to analyze what's working and refine their implementations of technology-supported innovations to iteratively achieve improvements in learning impacts.

Understand: What research tells us about the effectiveness of learning technologies

There are five overarching lessons from the research, each of which we will discuss in turn:

1. Finding disparate results for a given learning technology product is more common than not.
2. It's rarely possible to disentangle the impact of the learning technology from the effectiveness of the overall instructional system in which it is embedded.
3. Whether use of a digital learning product appears "effective" in a given setting depends on the purpose and goals that educators have for using technology in their classrooms.
4. Even learning technology products designed for independent use are experienced differently depending on the nature and level of support students receive.
5. When it comes to achieving learning impacts from a complex technology-enabled change in instruction, time will often be needed to iterate and learn from early experiences.

Finding disparate results for a given learning technology product is more common than not.

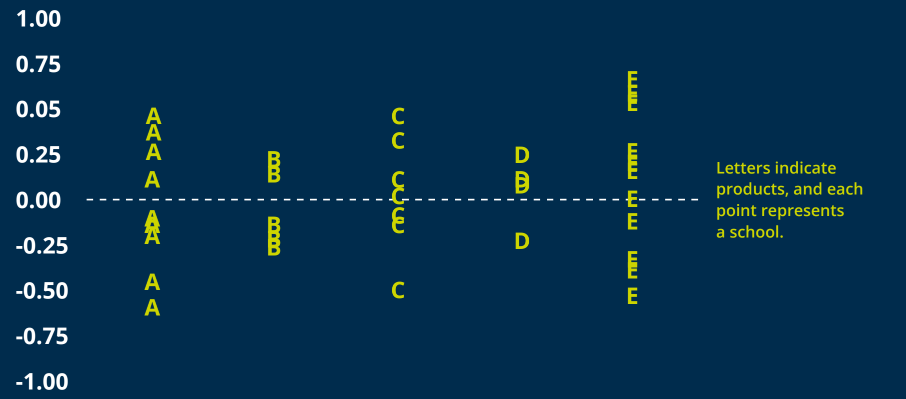
Given that a learning technology product can be used in very different ways in different classrooms and settings, we should not be surprised if learning gains associated with a product are substantial in some situations but small or even non-existent in others. Figure 2, for example, shows the estimates for school-level learning impacts for five first-grade reading software products included in a large-scale experimental evaluation of reading and math software.[8]

Impacts are shown as effect sizes, and those above the zero-line are cases where students in classrooms using the software had better reading achievement scores than students in other classrooms in the same school that did not use the software. Letters below the line represent cases where achievement was better for control classrooms.

Each letter (A, B, C, D, E) represents a different product. The important thing to note is that none of these products had consistently positive or consistently negative impacts across schools. Results for the other 11 reading and software products tested as part of the study were similarly diverse.

What Do We Mean By Effect Sizes?
An effect size is a measure of the difference in average performance of two groups in standard deviation units. An effect size of +0.5 would mean that the treatment group average was one half a standard deviation higher than the control group average, which is equivalent to scoring at the 69th percentile rather than the 50th percentile.

Fig. 2 School-Level Effect Sizes by Product, First Grade (SAT-9 Reading Score)



Source:
Dynarski et al., 2007.

Note: Statistical significance of average effect sizes cannot be inferred from the figure because a student and teacher sample sizes differ between schools.

It's rarely possible to disentangle the impact of the learning technology from the effectiveness of the overall instructional system in which it is embedded.
When we measure student outcomes resulting from blended models, we're really measuring the impact of a broader learning system. This will include an educator's instruction and actions – for example, explaining how activities within the learning software relate to other things the class has already learned or providing coaching for students as they use the learning technology. It will also include the interactions between students' past knowledge and the content of the learning software as well as the interactions between students and the learning technology per se.

If learning outcomes should be attributed to this broader constellation of educator and student activities as well as to the support that technology affords, we need to include those human activities in our conceptualization of what it is that proves to be effective or ineffective.

An example is provided by the Rocketship approach to instruction. Rocketship Education is a charter school network well known for its extensive use of learning technology and individualized student "playlists." Rocketship students spend a considerable portion of their school day in learning labs, working with online math and English-language arts software. In addition to the learning labs, other important components of Rocketship's instructional system include regular classroom instruction emphasizing small-group collaborative learning and extended school days. In addition, struggling learners, identified during weekly staff reviews of student performance data, receive daily one-on-one and small-group teacher-led tutoring sessions during lab time. When we studied Rocketship's personalized instruction model, we found that students in the Rocketship elementary schools participating in our research did quite well on California's state achievement tests.

The predominantly low-income students in these schools performed at levels similar to those of students in schools that serve affluent communities.[9] But, given the complexity of the Rocketship instructional model, it would not be appropriate to attribute these achievement outcomes solely to the software products used in the schools' learning labs. In theory, you could systematically strip away each aspect of the Rocketship model one by one to ascertain the point at which student achievement falls to more typical levels, but educators have little incentive to do so. The various components of the Rocketship model appear to work together as an instructional system that prepares students for state testing. The Rocketship model is better described as a technology-enabled instructional model rather than as a technology intervention.

Whether use of a digital learning product appears "effective" in a given setting depends on the purpose and goals that educators have for using technology in their classrooms.
Educators have many different goals for the technology they bring to their classrooms. They may view it as a way to introduce a spark of excitement to the learning process, individualize the content each student is exposed to, or enable flexible classroom management routines that enable them to work with small groups of students.[10] Before answering the question of whether a technology-supported intervention was effective, one must ask, "Effective at doing what?"

For example, in an SRI study of 23 implementations of nine adaptive learning software products in 14 colleges and universities, instructors in the treatment condition used these digital learning resources in many different ways and for quite different purposes.[11] In some courses the software was used as a practice environment for mathematics skills that had been taught in class; in some the software provided the core course curriculum; some used adaptive software tools to help students memorize the specialized vocabulary needed in the course; elsewhere the goal was to enhance students' independent study skills or to help instructors monitor student progress. Each college and instructor found a way to employ adaptive learning software, with implementation model choices influenced both by the college's institutional mission and by the instructors' instructional philosophy and conceptualization of the curriculum.

As a result, even when different instructors and students used the same adaptive learning software product, different impacts were observed across colleges and courses, just as seen in Figure 2 for first-grade reading software. Most instructors reported that the adaptive learning software they had worked with enhanced their instruction and their students' learning in some way, but those ways differed depending on the implementation model, and objective evidence of learning impacts was often lacking. Some of the colleges found evidence that student learning gains were larger with the software than in versions of the course without adaptive software, but most did not assemble the kind of data that would enable them to make this kind of determination.

When the purpose, context, and use pattern for the technology are this different in different classrooms, average learning gains across classrooms may provide little insight. To make sense of student outcomes, we must understand the intended outcomes and the actual roles of the learning technology in instruction in each implementation.

Even learning technology products designed for independent use are experienced differently depending on the nature and level of supports students receive.
Although the interdependence of the multiple aspects of instruction is perhaps most obvious in blended learning situations, we believe it pertains to fully online learning contexts as well. Students' success in online courses depends not only on the design of the course, but also on the types and quality of supports available to help students stay motivated and make adequate progress within the course. For example a study of 23 community college courses by the Community College Research Center found that the level of interpersonal interaction in a course was the strongest predictor of student grades. These supports are even more critical for the success of students who are struggling.[12]

Although the bulk of research on best practices in the implementation of online courses is observational and lacking in rigorous experimental designs, there is a growing consensus that certain scaffolds must be in place to ensure student success in online courses, especially for vulnerable learners.[13] Most successful online learning programs institute at least two layers of support:

- Remote monitoring and support by an online instructor, and
- Regular in-person check-ins with an instructor or counsellor responsible for the students' academic progress.

An instructor we interviewed during a visit to an online public university, for example, described how she leverages information she gains from the course software to tailor the supports she provides to her students. She goes into the instructor dashboard most days to see how both individual students and the class as a whole did with the software's practice exercises and formative assessments for the preceding day. She uses this information in deciding what concepts to focus on during her next synchronous online class and to recommend additional learning resources to individual students who are having difficulty with a certain problem type. She has even initiated Skype sessions with students to walk them, step by step, through problems they're having trouble with.

Some successful programs have added an additional layer of support at the pre-enrollment and enrollment stage by assessing students' readiness to take online courses, and providing individual counselling on course expectations and orientation programs to familiarize students with the instructional experience, navigation, and the courseware's various features including built-in learning supports. Researchers have found that attending an in-person orientation session for an online course is associated with a higher likelihood of succeeding in that course.[14]

Other programs have attempted to leverage the power of peer support by organizing regular opportunities for students enrolled in the same online courses in the same locality to physically meet with their peers. MOOC providers have been encouraging face-to-face "meet-ups" of students taking on of their massive online courses who live in close proximity to each other.

Absent a serious commitment to providing extra support for students enrolled in online courses, institutions and schools thinking about adopting online courses as part of their instructional offerings should not expect consistently successful outcomes, even when the course itself is well-designed and has been shown to be effective in supportive instructional settings.

When it comes to achieving learning impacts from a complex technology-enabled change in instruction, time will often be needed to iterate and learn from early experiences.
Instructional systems are complex and changes to these systems tend to happen slowly. The adoption of digital learning products often requires a fundamental change in the instructional environment and in the roles of instructors and students in the learning process. The greater the change in these roles for instructors and students, the more time will be needed for them to adopt and become proficient in their new roles.

As an example, we draw upon a recent study of Cognitive Tutor Algebra I—a blended learning curriculum that uses adaptive technology to simulate a human tutor and that requires teachers to adopt new ways of teaching.[15] The Cognitive Tutor curriculum involves students' use of the intelligent tutoring software for two days a week; three days are spent in complementary teacher-led collaborative learning activities. A two-year randomized controlled study of Cognitive Tutor Algebra involving 147 middle and high schools across seven states found positive effects on academic achievement for students in some classrooms assigned to use the curriculum. However, the results are nuanced. The researchers found significant impacts on achievement for students in high school but not middle school, and *only* after the second year of use by the teachers. The findings of this study highlight the importance of setting the proper expectations and taking a long-term approach to assessing the effectiveness of new adoptions of learning technologies.

Rather than being experienced immediately, the benefits of changing teachers' practices and linking adaptive technology-based practice with teacher-led collaborative learning may emerge over the course of multiple semesters or school years. The benefits will only occur after teachers and students adapt to their new roles and the necessary institutional supports for new practices are in place. The greater the change from standard practice required by the new approach, the longer it is likely to take for potential benefits to be manifested.

Implement: Using learning technologies effectively

Research shows that instructional innovations must be multi-faceted – that is, they must take into account the whole ecosystem in order to affect core teaching processes and learner outcomes. This means it will be necessary not only to identify an appropriate learning technology product or resource, but also to plan for the ways in which instructors and students will be changing their roles and routines to take into account the relationship between technology-based activities and other aspects of instruction. Below we consider four mutually dependent phases of this activity:

1. Identifying promising technology tools and resources that match the goals and context for the innovation,
2. Planning the multiple parts of the innovation as students will experience it and the supports needed for students and instructors to be able to implement the innovation as intended,
3. Implementing the multiple parts of the intervention as specified in the plan, and
4. Collecting and evaluating data to reveal how the innovation is being implemented and whether the innovation is having the desired impacts on student outcomes. Ideally, evaluation data are used to refine the technology implementation model (and sometimes the technology product itself) for future iterations of the intervention (i.e., cycling through the steps again, starting with Step 2).

Identify: Finding the right learning technology is itself a complex process

If it's unrealistic to expect to find a list of learning technologies that can be implemented in turnkey fashion and "work" every time, what do educators and education administrators need to do to maximize the likelihood that their investments in technology and technology-supported learning innovations will yield the results they're looking for? Below we present a series of questions that should be answered.

What is your goal for student learning?

Learning technology selection starts with understanding the student outcomes you're trying to achieve and the kinds of learning those outcomes entail. The best place to start is by articulating the nature of the learning challenge you're trying to address and how you would know whether students have made progress with respect to that challenge. This analysis has to be at a deeper level than "improving achievement" or "raising test scores." For example, your challenge might be:

- Helping students who are struggling with algebra because their basic math skills are weak
- Improving academic writing by instructing students in reading comprehension techniques or in analytic reasoning, depending on their individual needs
- Moving students who have learned how to compute a variety of different statistics to the next level where they know which statistical test to use when given a novel problem

Often, education institutions have some form of data they can use to help diagnose the source of student difficulties, but the time for analyzing and reflecting on such data and the expectation that instructors will do so are lacking. Opportunities for groups of educators teaching the same subject to spend time together reviewing student work and performance on assessments in detail can provide informed hunches regarding the sources of student difficulty.[16] In addition, the research literature on learning in specific content areas (e.g., mathematics, writing, physics) for specific kinds of learners (e.g., college, English learners, young children) can help provide guidance for this exploration.

Part of this goal clarification process is making sure you're clear about your priorities for student outcomes. It's difficult to optimize outcomes if you don't have a rationale for prioritizing them. Some kinds of learning experiences are best for helping students move quickly to retention of content in the short term, while different kinds of learning experiences produce superior ability to apply what has been learned in new contexts.[17] Mastery learning approaches, in which each student spends as much time on each learning objective as needed to reach a proficiency criterion, typically lead to higher scores on end-of-course examinations but lower rates of credit accumulation per academic term (because some students need more than a single term to reach the mastery criteria).[18]

What instructional designs support the learning goals you have targeted?

Once the most important learning goals motivating the adoption of technology have been clarified, the next step is classifying those goals in terms of the type of learning involved. Different instructional designs are conducive to different learning outcomes.[19] For example, direct instruction ("telling") is generally more efficient in enabling recall of factual content while exploration followed by guided problem solving is more conducive to knowledge transfer.[20] Acquiring skills such as arithmetic computation or fluent reading is facilitated by extensive practice on those skills with immediate feedback, as well as speeded practice (i.e., with time limits) at later stages of skill acquisition.[21] For acquiring conceptual understanding, on the other hand, techniques such as eliciting predictions and explanations are useful, as is starting instruction with an engaging, concrete real-world problem.[22]

A number of resources summarizing learning research can support efforts to identify the most appropriate instructional designs.[23]

Are there technology products that offer advantages for providing conditions that support this kind of learning?

Digital technology can support various kinds of learning. Technology can support the extensive practice with immediate feedback that is important for skill acquisition, for example, because it can offer an unlimited number of practice trials, generate immediate feedback after each response, and adjust difficulty level to the individual learner. To support learning factual content in a way that leads students not just to recall isolated facts but to build an understanding of a domain such as biology or geography, digital technology can provide an engaging real-world narrative or problem context with prompts for making predictions and giving explanations. Arguably, skilled instructors provide these kinds of conditions without using technology, but it is nearly impossible for an instructor to provide the ideal amount of practice with immediate feedback or to elicit thoughtful student predictions and explanations from every student in a large class. Technology provides scale. Potential adopters of a digital learning product should become familiar enough with the product being considered to know whether the experiences the product provides incorporate techniques known to support the kind of learning that is the goal of the technology adoption.

How good is the evidence of this product's impact on learning?

After reviewing learning technology products and finding one or more that are a good match for the kinds of learning you've targeted, it's prudent to look for evidence that use of the product has enhanced this kind of learning in the past. Reports of impacts may be available from curated independent repositories such as the U.S. Department of Education's What Works Clearinghouse, from study reports in academic journals, or from journalistic accounts in newspapers or trade publications. Increasingly, education companies are also releasing reports containing evidence of product impact.*

Whether they come from a journalist's account, an academic paper, or a government resource, conclusions or ratings about a technology-supported innovation's impact or lack thereof should not be taken at face value. It's important to delve more deeply into the basis for the impact claim to determine whether it's justified and relevant to your own situation. Sorting through impact evidence claims can be confusing. We suggest several broad criteria for deciding how much weight to give a report regarding impact:

- *Strength of the study design.* The design should rule out alternative explanations for improved outcomes, including student maturation, other changes being made at the same time the technology-supported innovation was introduced, and selection bias. (These design considerations are discussed further in the next section below.)
- *Outcome relevance.* If your goal for making a major change in instruction is to prepare students better to apply what they learn in future courses or work, research studies that measure only immediate learning impacts won't tell you what you want to know.
- *Implementation similarity.* Study results will be informative to the extent that the treatment group used the technology product in ways comparable to those you are planning in terms of duration, educator support, and ancillary activities
- *Contextual relevance.* Every learning situation is unique in some way, but there are also similarities that cut across contexts and provide grounds for expecting some similarity in outcomes if the intervention is implemented in a comparable way. The degree of confidence we can have that impact findings will generalize will be larger to the extent that the students, instructors and settings in the research study are similar to those in your context.
- *Objectivity.* A study conducted by the individuals or organization that developed the technology is more likely to report favorable results than one conducted by an independent third party.

As discussed above, disparate findings with respect to learning impacts in studies involving a learning technology product are the norm rather than the exception. Learning technology consumers need to identify the evidence most relevant to their own situation and to look at the preponderance of relevant findings rather than expecting a definitive "thumbs up" or "thumbs down" from the research literature. For newer products, of course, this will be much more difficult as there's not likely to be much evidence regarding impact immediately available. In such cases, there is an especially strong rationale for trying out the use of the product in a few classrooms or schools to generate preliminary impact evidence prior to implementation at scale, an idea we discuss further later on in this report.

How well does this product match to circumstances in the settings where it will be used?

Technology selection needs to include a frank appraisal of the technical and human infrastructures for software implementation. It's obvious that a Web-based product that requires considerable bandwidth places demands on the network connections of education institutions and classrooms.[24] Planners should check that there's an adequate supply of functioning hardware that meets the product's software specifications. Any needed upgrades to the technology infrastructure and technical support capabilities should be in place before instructors are asked to use the technology product with their students.

Plan: Getting consistently positive impacts from the use of learning technology requires attending to the multiple aspects of the instructional system

We have made the case that learning technologies should be thought of as one component of a broader instruction system or innovation. For this reason, preparation for using a new technology requires planning out the whole system, not just making a technology purchase decision. In our research, we have identified five essential pieces for an effective learning technology implementation that should be covered in the implementation plan:

- Leadership support;
- A technology infrastructure that can supply adequate and reliable access to the technology without undue dropped connections, system crashes, or unacceptable load times;
- Time within the schedule for all students to receive the recommended "dose" of learning technology use;

*** For example, Pearson, which supported this work, has been rolling out reports related to their products' efficacy since 2011, and has committed to releasing independently audited efficacy reports on it products by 2018.**

- Alignment between learning materials and activities, technology-based or otherwise; and
- Role articulation, training, and ongoing support for educators.

If any of these elements of implementation is not sufficiently in place, the likelihood of a significant positive impact declines.

Leadership support is important because education leaders are in a position to garner needed supports for the innovation, such as time for instructors to be trained on how best to use it or any needed equipment. They also are in charge of other policies and practices which may either support or hinder instructors' ability to implement the technology-supported innovation. The required technology infrastructure is an obvious prerequisite, but we have found that individuals making decisions about learning technology adoptions are sometimes unaware of issues such as the number of school computers that are inoperable or the impacts that district firewalls or simultaneous use of the Internet in a large number of classrooms can have on access to online resources.

Time requirements are another straight forward prerequisite for technology-supported innovations that often gets overlooked. Planners should figure out how much time the anticipated use of the learning technology will require, including time for shifting to technology use (which may require students to move from classrooms to computer labs or to check out laptop computers). These requirements then need to be considered in light of other activities and innovations the same instructors and students are expected to execute. A 50-minute grade 7 English class in which teachers are required to have students do 15 minutes of silent sustained reading every day is not going to be able to implement writing instruction software with activities that can't be completed in less than 40 minutes, for example.

A more subtle but equally important consideration is alignment between the content and instructional philosophy behind the learning software and those of other instructional resources students will use and the methods employed by their instructors. Incompatibilities can occur in the way in which content is presented in the software and in the instructor-led portions of a blended course. For example, mathematical operations on positive and negative rational numbers can be introduced using a number line or through teaching a set of rules for the order of operations in a linear equation. If the students' textbook uses one approach and their learning software uses another, they may not even realize that they're doing the same thing in both contexts. Students do better when they see links between their learning experiences with the software product and other things they're doing in class. Educators can help identify these links, but only if they themselves are familiar with the software.

Finally, the implementation plan should identify who will do what in terms of preparing for and implementing the technology-supported innovation. Teachers will do better incorporating learning technologies into their classes if they have received training on those technologies—not just the mechanics of using the software, but recommended instructional practices as well. It's also important for planners to take into account the amount of time instructors will need to spend preparing for technology implementation (for example, entering class rosters or their own content into the system). In general, the bigger the change required to existing practices and organizational support systems, the more time and support implementing a new learning technology is likely to require.[25]

Execute: Implement the plan and track progress

When a technology-supported innovation is a major investment, education leaders are often tempted to evaluate impacts after just one semester or year, trying to demonstrate success quickly to their stakeholders. Often, these early results are disappointing.

This timeline for evaluating impacts is not consistent with the extended process we have described for supporting fundamental changes to teaching and learning practices. Moreover, there's no point in trying to evaluate the effectiveness of an instructional system that was never really implemented as designed Unfortunately, the evaluation literature is replete with examples of learning technology initiatives with poor or limited implementation. For example, in 2001 a large American school district spent over $60 million on a combined hardware and software product to teach early reading skills.[26] The software was intended as a supplement to the core reading program for kindergarten and grade 1. Implementation of the reading software had to compete with that of the mandated core curriculum. System data later showed that kindergartners got less than half of the time with the software they were supposed to and first graders got less than a third of the recommended time.[27]

It's important to be aware and make allowances for school or district policies that are incompatible with the intended implementation practices for the new technology. In the case of the early reading system implementation described above, evaluators examined the district-mandated time for literacy instruction, class sizes, and the number of computers in classrooms. They found that in many classrooms it would have been literally impossible for all students to get the recommended time using the learning software during the available literacy instruction block.[28] Such incompatibilities should have been identified in a planning stage like the one we described above. But in any case, it is beneficial to couple execution of the implementation plan with measurement of the extent to which critical elements of the plan are being put in place in each participating school, department, or classroom.

Ascertaining the extent to which these features are present in each of the settings where the learning technology undergoing evaluation is being used is good practice. If resources permit, collecting quantitative measures of these aspects of implementation (for example, through system log data and structured observations or surveys) can be useful during implementation to spot classrooms and schools where troubleshooting is called for. The data also can be useful after implementation to help make sense of observed differences in learning impacts.

Evaluate: Use data to make judgments about the technology-supported innovation

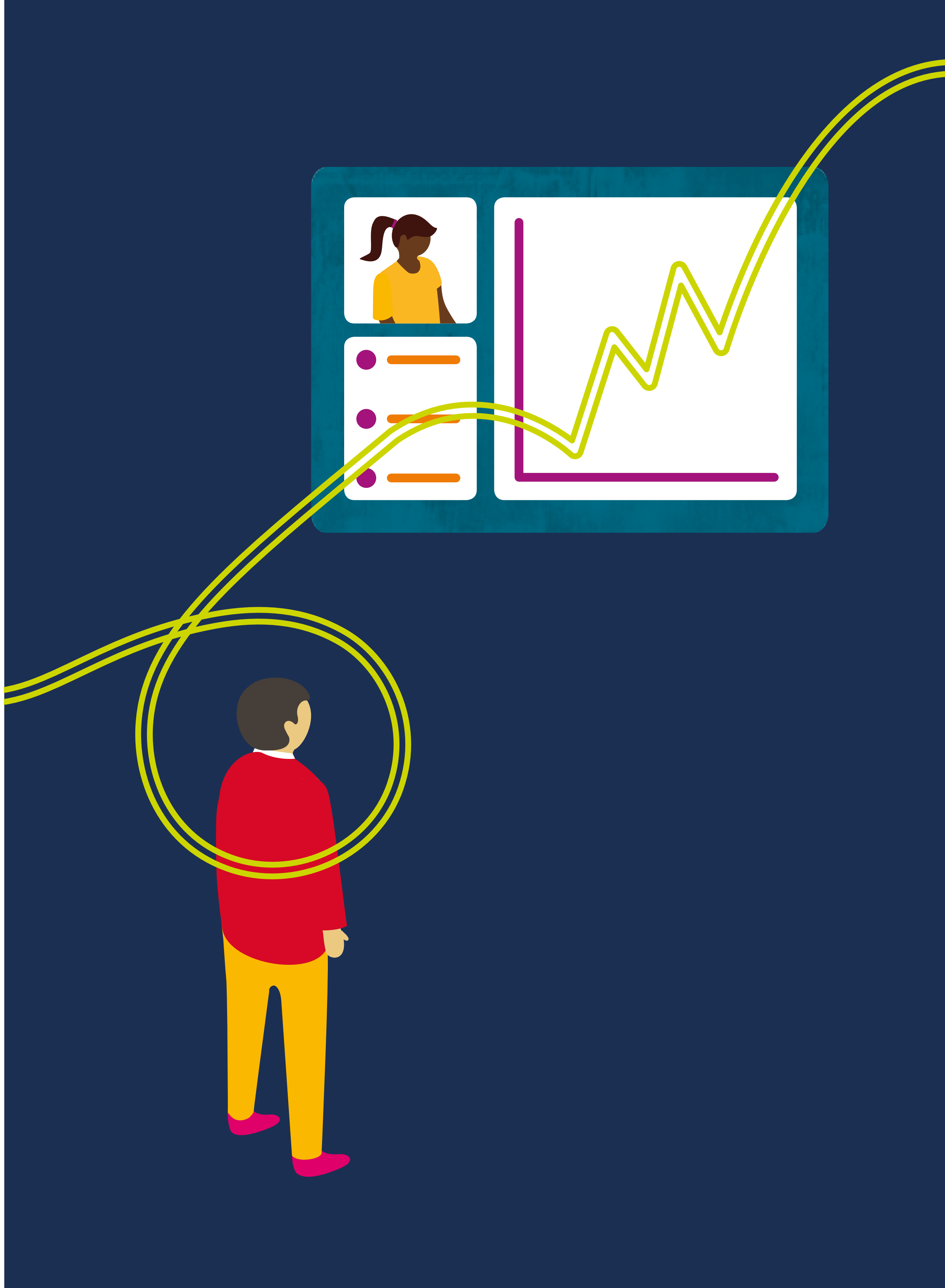
Because the results of a technology-supported innovation typically vary from context to context, an institution adopting it should evaluate its success as they have implemented it with their own students. The remainder of this brief will discuss a number of evaluation options at some length, but here we note three prerequisites that apply to all of them:

- *Goals.* Being clear about the key outcome or outcomes the innovation is targeting (for example, to improve students' ability to critically analyze the arguments made in historical texts) is the first prerequisite. You won't be able to measure progress if you're not sure what outcome you're trying to attain with the technology-supported innovation.

- *Measures.* Beyond knowing the instructional goal, decision makers need to figure out how they'll measure whether or not the instructional system achieves that goal. Hopefully, the software itself incorporates learning assessments and provides data on student performance on these assessments over time. But we recommend supplementing any measures internal to the learning system with a measure of the target skill outside of the learning software product itself. In the case of the example of critical analysis of historical texts, a compare-and-contrast essay assignment that could be administered both to students who have used the software and to students who have practiced this skill through other kinds of activities could serve as the external outcome measure.

- *Research questions.* Individuals involved with the technology-supported innovation are likely to have many questions about how and how well it works. The may want to measure its impact, which requires comparing outcomes for students experiencing the innovation with those of comparable students experiencing some alternative form of instruction. Alternatively, they may be more interested in understanding which components of the innovation were difficult to implement and why they were difficult. There are never time and resources to answer all possible research questions, and decision makers need to set priorities.

If a major investment is to be made in a rigorous study of impact, then, it should be timed to coincide with program designers' expectations of how long it will take for strong results to be demonstrated. This does not, however, suggest that all research should be delayed until after the initiative has matured. Instead, research methods should be selected to align with the stage of implementation they are meant to inform. As educators are beginning to try out learning technology tools and experiment with new opportunities for instruction, methods that are more deliberately informative to ongoing practice can be a powerful support to evolving success, as we will describe below.



Evaluate: Choosing the right method

An Improvement Science Approach to Evaluation

Although many people think of evaluation as a summative judgment of efficacy, it can also be an essential part of continuous improvement efforts. A fifth phase can be added to the four described in the last section: Refine the technology-supported intervention by using evaluation data to improve the technology and/or implementation practices.

Anthony Bryk, Louis Gomez and colleagues describe such an approach.[29] "Improvement science" teams pair educators with researchers to conduct a series of short, targeted studies, each with a well-defined goal appropriate to the stage of adoption. These iterative cycles use research to inform and support the change process for instructors and systems. Instead of asking about impact ("How well did it work?"), improvement science asks "What is the next challenge we need to solve?" The aim of improvement science studies is to refine implementation practices in ways that lead to better outcomes than obtained on the last iteration. These studies are not intended to provide the kind of evidence that would help you decide whether or not a learning technology is worth implementing. But they can be used to help you decide how to implement it better. The developmental mathematics Pathways work led by the Carnegie Foundation for the Advancement of Teaching provides a concrete illustration of this approach. (see "Measuring Efficacy" case study on next page).

For research that supports a continuous improvement process for implementation of a technology-supported innovation, a number of considerations are important:

- *Focus on the important problem to be solved.* Is this a program to assess low literacy levels? Is summer school experiencing low success rates and high cost? Each improvement cycle is conceived as a step toward a clearly stated, measurable long-term goal
- *Attend to leading indicators.* Of course it is important to have a clear and agreed-upon measure of the long-term goal (e.g., test scores or student retention rates) from the beginning of the initiative. But because those long-term successes may take some time to emerge, it is important to describe and track initial and incremental outcomes. If we adopt a blended learning program, we might not see test scores improve in the first semester, for example, but a more immediate outcome could be the changes in teaching practices the program is supposed to catalyze.
- *Success requires more than the software*. As we have described, other essential ingredients of implementation include articulation of the new practices expected of educators and provision of supports for educators to learn them. One or more improvement cycles might, for example, attend to the design of the professional development that instructors receive with the initiative or the removal of barriers to adoption of those new practices.
- *Consider the use of system data*. Many learning technology systems provide a host of data about student learning paths and behaviors as well as tracking outcomes. Thoughtful incorporation of these data into improvement cycles can help instructors understand whether some students are not engaging with the learning software frequently enough to attain their learning goals or whether they are interacting with the software in productive or nonproductive ways. System data can reveal, for example, how much time different students spend trying to work out a problem before they ask the learning system for a hint. Such data can inform not only immediate instructional decisions but also the design of human supports that can help students engage more productively.

Pathways: An Improvement Science Approach to Evaluation

In American community colleges, the typical "developmental mathematics" courses that more than half of community college students must take are notorious barriers to degree attainment for students who have not previously qualified for college-level mathematics. While estimates vary, one study of 57 colleges found that only 31% of students who were assigned to a sequence of developmental mathematics courses successfully completed that sequence within three years.[30] With such poor course success rates, developmental mathematics is a major obstacle to the postsecondary progress of vast numbers of students.

To address these egregious concerns, a networked improvement community was formed in 2010 among researchers, led by the Carnegie Foundation for the Advancement of Teaching and the Charles A. Dana Center at the University of Texas, and educators from 28 community colleges. This improvement community engaged in iterative cycles of research, development, and refinement to develop two blended courses, Statway® and Quantway®, intended to support students through a developmental mathematics trajectory in a single year.

As a first step, the collaborating institutions worked on unpacking the problem: what was holding students back from graduating that could then become specific targets for improvement? Based on prior research, the team identified students' doubts about themselves as able mathematics students as one of several primary inhibitors of success for the population of students enrolled in developmental mathematics courses, and a focus on "productive persistence" as one of the frameworks that would guide program design and instructor interactions with students. As a result, the program featured deliberate treatment of the psychological aspects of engagement in mathematics, such as the relationship between intelligence and effort, following the work of Carol Dweck.

This hypothesis was then tested in a number of rapid classroom-based experiments at the various participating community colleges. For example, in one study students who read an article about the ability of the adult brain to learn based on effort were twice as likely as peers in a comparison group to complete the course; these students on average also achieved a significantly higher grade point average.

Targeted and tested design elements such as this seem to have contributed to a successful program. Overall, 56% of Quantway students completed their developmental mathematics requirement and earned a college mathematics credit in just the program's first semester, compared to 21% of their same-college peers in different developmental mathematics programs that completed the requirement in a full year. Statway students were similarly dramatically more likely than past developmental mathematics students at the same institutions to have earned a college mathematics credit within a year of continuous enrollment.[31]

Evaluating Impacts

More Info
For readers interested in additional information about rigorous evaluation design, there are many comprehensive resources available.[32]

The approach of iterative short-cycle research described above is appropriate for refining implementation models over time. In some cases, that kind of small-scale work integrated into ongoing practice will be all that is required. But in the case of large-scale learning technology innovations, there is often a need for summative reporting to stakeholders ("the $2 million in taxpayer money invested in new computers, software, and wireless access resulted in . . . ."). Rigorous studies of impact are appropriate also in advance of high-risk decisions to take a new instructional system to very large scale or to make it mandatory.

Below, we introduce several types of impact studies that are commonly discussed in connection with learning technology, and offer some pointers to important design features.

Credible impact research requires use of a comparison or control group.

Examining outcome measures (such as course grades or scores on a final examination) is a necessary but not sufficient aspect of evaluating impact. Looking at these measures solely in classes that are using a particular online or blended learning model does not help to understand how much of the learning that students exhibited after experiencing the technology-supported innovation can be attributed to that innovation. Presumably, students would still have learned something if they were in a class on the same topic that used different methods and tools. Is a 10% increase on an assessment from the beginning to the end of the semester more or less than these same students would otherwise have achieved?

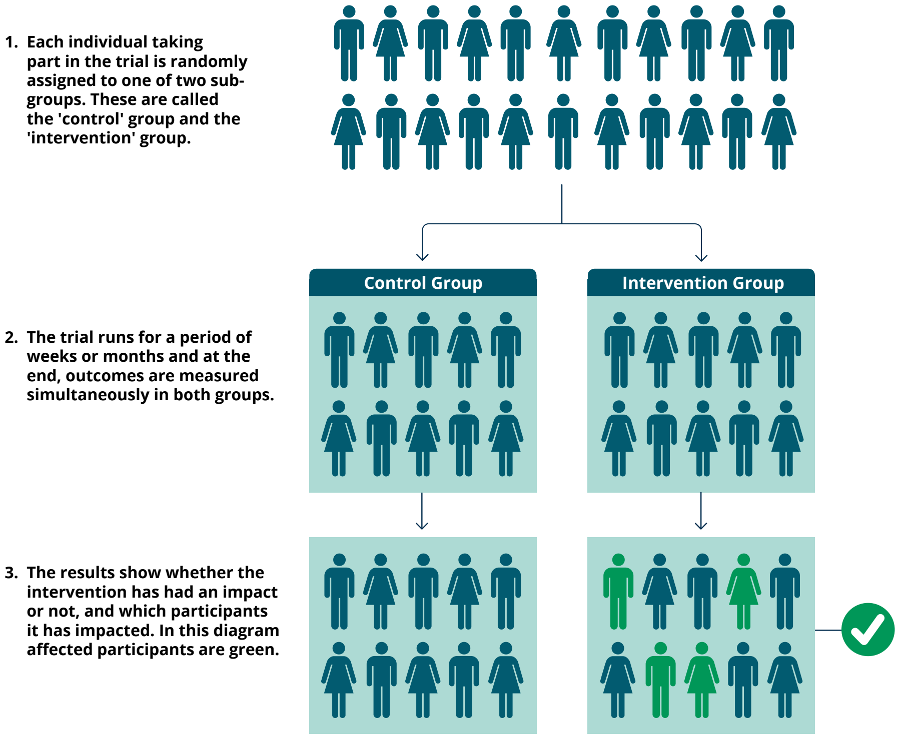
For this reason, impact studies compare student outcomes in *treatment* classes (that are experiencing the new program involving learning technology) to those in comparison or control classes that represent the type of instruction students would have been exposed to if the program had not been available. The performance of students in these latter classrooms is intended to approximate the types of outcomes that treatment students would have achieved with "business as usual" instruction. The challenge addressed by rigorous impact research design is that, because it is not possible to turn back the clock for these same students, the outcomes for different students in a control condition serve as proxies for what the outcomes of the treatment group students would have been without the new program.

The strongest method for establishing equivalence between treatment and comparison groups is to randomly assign students, instructors, or schools to treatment and comparison conditions.

True experiments (sometimes called randomized controlled trials or RCTs) produce stronger evidence of causation because random assignment can be presumed, on average across a large enough number of cases, to eliminate selection bias. That is, if each of 300 students signed up for introductory biology is assigned by chance to either experience the new technology-supported version of the course or to experience the traditional version, we can assume that on average the two course versions will have pretty much the same proportion of females, high-achieving students, English language learners, and so on. (Although it's still a good idea to check that the random assignment achieved this outcome.) To support high-risk decisions, where the consequences of a bad learning technology choice would be very serious for students and there is little prior evidentiary backing for the technology-supported program, RCTs are the preferred method.

However, large-scale random-assignment experiments are not always practical because they can be costly (in terms of both resources and time) and require conditions that can be challenging to arrange. For example, random assignment of students to course sections in which much of the learning is done online or to traditional classroom-based course sections can be unacceptable to institutions wary of forcing students to learn in an online environment if they don't wish to do so. In these cases, there are alternative approaches.

Fig. 3 How does a randomized control trial work?



Alternatives to random-assignment experiments can be credible if they do a good job of demonstrating the equivalence of treatment and comparison groups before the implementation.
Where randomized experiments are not practical or desirable, a well-designed quasi-experiment is a common alternative that can provide reasonably compelling evidence of effectiveness. Instead of randomization to insure the treatment and control groups were equivalent before the former group was exposed to the new instructional system, these methods use pre-tests (measures of the competency that the initiative is trying to improve) or predictive data from student-level administrative records (such as student demographic characteristics and prior achievement) to ascertain to the extent possible whether students in the treatment and comparison groups were equivalent at the outset. To the extent that the two groups were equivalent prior to instruction in terms of all the variables that might influence the learning outcome, a case can be made that differences in outcomes can be attributed to the new technology-supported instructional system. Where there are pre-existing differences between the two groups, these may be controlled for statistically, provided they are not too great in magnitude.[33]

In designing quasi-experiments, it is important to do as much as possible to mitigate selection bias: that is, pre-existing differences between the two groups of students and instructors being compared. Quasi-experiments contrasting volunteers (students, instructors, or schools) with nonvolunteers are always open to criticism. Even if the two groups appear similar in terms of their backgrounds and prior achievement, questions may arise as to whether the teachers who sign up for the new program are instructional innovators in other respects, or whether students who choose a technology-supported version of a course are more or less motivated learners than their peers.

Research designs without an equivalent comparison group or an objective learning measure cannot provide credible evidence of impact.
Less costly but less rigorous research designs include measuring gains from a pretest to a post-test for students experiencing the technology-supported program only or collecting student and instructor satisfaction ratings and perceptions of how much they learned. The problem with the first of these approaches is that we would expect students to learn something from the beginning to the end of the course, and without collecting data on a comparison group we can't tell whether or not the technology-supported program is an improvement on standard practice. The satisfaction rating approach falls short because such ratings do not have a strong relationship to learning impacts.[34] Nevertheless, instructor and student satisfaction are important considerations in their own right, and in low-risk cases such as the implementation of a technology product as a supplemental resource rather than a major part of the course, or as part of the more formative stage of research described earlier, they can be useful.

The Advantages of Starting Small

Researchers who have studied improvement processes in schools and colleges advise "starting small and learning fast."[35] In other words, it is usually desirable to introduce a new instructional system on a small scale and measure its results before moving to wide-scale implementation. This insight is just as relevant to the introduction of new learning technologies as to other educational innovations. By starting small, an education system can confine any negative consequences of introducing the innovation—either poor outcomes or unanticipated side effects—to a small number of students and instructors. Further, data can be collected from the small-scale trial and used to refine the new instructional system before it is implemented more broadly. In addition, the organization and staff members involved in the small-scale trial will develop insights and expertise that can benefit others when the innovation is rolled out more broadly.

The University of Maryland University College's approach to trying out new software for its Introductory Statistics course demonstrates this approach. UMUC offers instruction entirely online in the form of eight-week courses. Roughly 13,000 students take Introductory Statistics each year from as many as 75 different online instructors. When UMUC became interested in trying out the Online Learning Initiative (OLI) Statistics course, it did not adopt the courseware for all its statistics classes. Rather, it had just three instructors use the courseware in a handful of course sections so that it could examine student outcomes with OLI Statistics and compare them to those of students in other sections of the course. One of the instructors for this small-scale test, who had previous experience using OLI Statistics in another course, was able to prepare materials on how to use the courseware that she shared with the other two instructors in the pilot.

If UMUC does go forward with a systemwide adoption of OLI Statistics, it will go forward with insights developed from the pilot study and with three instructors experienced with using the software who can help mentor other statistics faculty on how to use the learning technology to best advantage.

Rapid Lightweight Learning Technology Evaluation Approaches

The speed with which blended learning models tend to evolve can be a challenge for lengthier methods that rely on well-defined, steady conditions. Because of these and other specific requirements, experiments should be undertaken with careful consideration of the institution's implementation readiness, available resources, and practicalities. Sometimes experiments simply won't be appropriate, and other, more "lightweight" experiments can be used.

The software industry often employs a kind of random-assignment experiment known as A/B testing to compare two different versions of the same technology product (version A and version B) by randomly assigning users to one or the other version. Historically, A/B testing has its roots in market research, such as for comparing the sales or click-through results of two user interface designs or two versions of an advertisement. But increasingly it is being applied to digital learning research and development, and online learning resources that attract many users can run A/B tests comparing alternative versions in a short amount of time.

The Khan Academy, for example, reports that it attracts enough users to run an adequately powered A/B test in a matter of hours. In A/B tests involving smaller numbers of users, user characteristics and prior achievement matter more, but rapid experiments are still possible. At the Center for Advanced Technology in Schools at the University of California, Los Angeles, for instance, researchers ran 20 randomized controlled trials over an 18-month period to test various theory-driven hypotheses about learning game design.[36]

Providers of popular learning platforms have additional options to harness technology for conducting rapid tests of learning impacts. A pioneer in this field has been PowerMyLearning, a nonprofit organization based in New York City. The free PowerMyLearning Connect platform hosts scores of learning applications from multiple sources (such as Khan Academy, BBC, Starfall, and PBS) keyed to specific learning objectives. Seeking a cost-effective way to find out which learning applications are most effective, PowerMyLearning began setting up rapid online experiments in 2014. Each experiment is a “horse race” between two applications targeting the same learning outcome. When students enter a “Mission Module” on PowerMyLearning Connect, they are assigned at random to experience digital learning application A, to experience digital learning application B, or to simply proceed to the post-test. Since a large number of students are randomly assigned to condition, it is assumed that their average achievement level prior to the experiment is equivalent and average post-test scores can be compared between students experiencing learning application A and those experiencing learning application B.

Another approach has been championed by the Office of Educational Technology (OET) within the U.S. Department of Education. OET wanted to make it possible for school districts to conduct rapid, low-cost evaluations of learning technology products. The strategy they are trying out involves linking learning system use data to the longitudinal student data records that districts maintain. In this way, districts can see whether students exposed to a particular learning software product perform differently on district and state achievement tests than do students who do not use the product. In carrying out this work for OET, the research firm Mathematica Policy Research has suggested that school districts could use a lottery to determine which schools, classrooms, or students get to use a new learning technology in its first year, in effect creating a random-assignment experiment. Alternatively, every student or teacher in the district could be given access to the new learning technology, but only a subset chosen at random would receive encouragement or “nudges” to use the technology, creating something like a randomized controlled trial assessing the impact of different product usage levels.

Self-Assessing the Rigor of Your Planned Summative Evaluation

Education systems seeking a rigorous design for summative evaluation of their technology-supported intervention should ask themselves whether their design meets some basic requirements:

- *Does the research compare treatment and comparison or control groups?* Often, the comparison will be between classes that use the software and a *“business as usual”* condition. Alternatively, two different innovations can be compared
- *Is the sample size large enough to be meaningful?* The smaller the sample, the more likely it is that results for one or two extreme individuals or classes will affect the average unduly. Moreover, the larger the sample, the smaller the effect your study can detect. Except when students are assigned to treatment and control groups, the total number of students participating in the research is not the critical aspect of sample size. If teachers are recruited for the study and assigned to conditions, it is the number of teachers. If schools are assigned to conditions, it is the number of schools. The size of sample you need depends also on factors such as the amount of diversity in the population you want to generalize to and the size of impact you’re interested in measuring.
- *Is there a common performance measure for both treatment and comparison groups?* If software-based measures are used for outcomes in the treatment group, often no equivalent measure is available for the business-as-usual classroom, precluding a comparison of outcomes.
- *Is there a common measure of proficiency prior to instruction?* A pretest measure is the best control for pre-existing differences that influence the post-test. With both pre- and post-test scores, researchers can look at gains over a defined period of time. If a pretest on the same content covered in the post-test is not available, it is still useful to use another measure of prior achievement (such as score on the prior year’s state achievement test or grade point average) in the analysis.

Comparison of outcomes across conditions can tell you if the use of a technology-supported instructional system had an impact on learner outcomes, but not how or under what conditions. Because effectiveness is a result not just of a learning technology but also of other components of the instructional system within which it's used, it is important to measure and report on these components as well as outcomes.

If implementation processes were carefully planned, recorded, and monitored to inform designs and refinement of teaching practices, as recommended above, you should already have multiple measures of implementation. For rigorous summative research to have meaning, the contrasting conditions for which outcomes are measured should be described with at least the types of information on the groups being compared described in the table below.

Criteria	Definition	Comments
Role of the learning software.	Aspects of the course provided through the software and those provided by a human instructor. The software could have served as the primary source of core curriculum content, provided a practice or formative assessment environment, or offered supplemental or enrichment content.	Course outcomes and software usage data need to be interpreted in the context of the role intended for the software.
Usage of learning software.	Amount of time learners spent using the software. May be measured in minutes, number of sessions, or number of completed modules.	If students only experience a learning technology for an hour a week, learning gains are harder to attribute to the software than if students used it every day.
Pedagogy in treatment and comparison classes.	The instructional approach used most often. Could be direct instruction (telling), skills practice, inquiry/exploration, or collaborative knowledge building.	Typically, the potential for improved outcomes is greater when the adoption of learning technologies is used as an opportunity for redesigning an entire course using learning science principles.
Retention rates for treatment and comparison groups.	Proportion of the students who started the learning experience remaining in it until completion or to the point when the post-test was given.	If more lower-performing students drop out of the online course than the traditional course, class-average learning results will be artificially inflated for the treatment group.

The Piñata Game: When System Data Isn't Enough

Internal system data can provide useful measures of implementation and student behaviors when interacting with particular designs. Internal data, however, can be misleading when viewed on its own. It can be powerful to pair appropriate use of internal system data with off-platform measures of context and use as well as external performance measures. Early development of an award-winning series of preschool games provides an example.

The Piñata Game was an early version of a game idea developed by content developers from the WGBH television station in Boston, Massachusetts, working with a team of researchers from EDC and SRI International. The game was intended to be an engaging opportunity for students to practice the core early math skill of "subitizing": the ability to recognize the number of objects in a small set (for example, to recognize that there are 3 objects in a group). In an early prototype pairs of students were shown a small number of items, and challenged to "catch" with a virtual blanket any group of items falling from a piñata that matched the target set in terms of number (e.g., any group of 3).

In testing this early version of the game, members of the R&D team noticed some pairs where one child anchored his or her index finger on one end of the screen and all the blanket movement was created by the other child, who moved his or her finger back and forth to catch objects or let them drop on the side of blanket. If the technology developers had looked only at the data captured automatically by the game, it would have appeared that the child who just anchored a finger was loafing and letting the other child do all the work.

But because the team was actually observing and videorecording game play, they could see that, while this was sometimes the case, in other instances the child who was not moving his or her finger was actually doing most of the mathematical thinking. As each set of objects fell out of the piñata, this child would say "Get it! Get it!" or "No, no, no!" while the other child handled the physical requirements of the task. In this case system log data would have looked identical (one child holding still, the other actively catching items) for two very different forms of student interaction: one in which the child who held still was disengaged, and the other where the child who held still was not only highly engaged, but was actually in charge of the mathematical thinking.

Conclusion

It has become increasingly clear that the question is no longer whether technology will be used within education but rather how best to leverage digital technologies to enliven education, enhance student outcomes, and make education practices more efficient. These goals are attainable, but not by pressing a button. Much of the rhetoric around learning technology effectiveness—both pro and con—suffers from a basic mischaracterization of the technology product per se as an intervention. We have argued that the intervention always includes more than just the technology components, and therefore that it doesn't make sense to talk about a "proven effective" learning technology product outside the context of its use. It is important to understand the possible implementation models and support features that in combination with the technology product can achieve desired outcomes. We should expect getting consistently positive impacts from a technology-supported initiative to be challenging. But that challenge can be surmounted if we plan initial implementations carefully, try them out first on a small scale, learn from multiple iterations and refine the implementation model and support system as we go.

We do not wish to leave readers with the perspective that all digital technology products are created equal and it's only implementation that matters. There are fundamental findings about learning that have been replicated over and over. Technology products based in this research are more likely to have positive outcomes. We know, for example, that learning experiences that require students to actively process conceptual content and relate it to what they already know will produce more lasting learning than will just putting factual information online for students to read.

The approach we present here entails establishing a discipline of focusing on student learning outcomes and accumulating evidence and using it in multiple iterations. It requires thinking systemically about desired learning outcomes and about student, instructor, and technology roles in instruction that can produce those outcomes. It calls for measuring as you go and for making mid-course corrections. While the scope will differ, the basic principles are the same whether applying this approach to the individual classroom, the school, or an entire education system. The approach we've advocated is much more complex than picking a particular product and making a large-scale purchase, but we hope education systems will consider this approach not as a burden but as a golden opportunity to improve their effectiveness by leveraging technology.

References

1. Bransford, J., Brown, A., & Cocking, R. (1999.) *How people learn: Brain, mind, experience, and school.* Washington, DC National Research Council.

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

3. Adkins, S.S. (2017). *The 2016 Global Learning Technology Investment Patterns Analysis.* www.metaari.com/whitepapers.html

4. Cuban, L. (2001). *Oversold and underused: Computers in the classroom.* Cambridge, MA: Harvard University Press. Chafkin, M. (2013 November). Udacity's Sebastian Thrun, godfather of free online education, changes course. *Fast Company*, ; Federal Communications Commission. (2016). *2016 broadband progress report.* Washington, DC: Author.

5. OECD. (2015). *Students, Computers and Learning: Making the Connection.* Paris: OECD Publishing.

6. Fullan, M. and Donnelly K. (2013). *Alive in the Swamp: Assessing Digital Innovations in Education.* London: Nesta; Luckin, R., Holmes, W., Griffiths, M. and Forcier, L.B. (2016). *Intelligence Unleashed: An Argument for AI in Education.* London: Pearson.

7. Cohen, D., Raudenbush, S., & Ball, D. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 1-24.

8. Dynarski, M., Agodini, R., Heavisude, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H., Emery, D., & Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort. Report to Congress. NCEE 2007-4006.* Washington, DC: U.S. Department of Education.

9. Murphy, R., Snow. E., Mislevy, J., Gallagher, L., Krumm, A ., & Wei, X. (2014). *Blended learning report.* Austin, TX: Michael & Susan Dell Foundation.

10. Murphy, R., Gallagher, L., Krumm, A .,Mislevy, J., & Hafter, A. (2014). *Research on the use of Khan Academy in schools.* Menlo Park, CA: SRI Education.

11. Yarnall, L., Means, B., & Wetzel, T. (2016). *Lessons learned from early implementations of adaptive courseware. Menlo Park, CA: SRI Education.*

12. Community College Research Center. (2013). *Creating an effective online instructor presence.* New York: Teachers College, Columbia University.

13. Bakia, M., Anderson, K., Heying, E., Keating, K., Mislevy, J. (2011). *Implementing online learning labs in schools and districts: Lessons From Miami-Dade's first year.* Menlo Park, CA: SRI International. Bakia, M., Mislevy, J., Heying, E., Patton, C., Singleton, C., & Krumm, A. (2013) S*upporting K-12 students in online learning: A review of online Algebra I courses.* Menlo Park, CA: SRI International.

14. Wojciechowski, A., & Bierlein Palmer, L. (2005). Individual student characteristics: Can any be predictors of successes in online classes? *Online Journal of Distance Learning Administration, 8.*

15. Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale. *Educational Evaluation and Policy Analysis, 36*(2), 127-144.

16. Mandinach, E. B., & Gummer, E. S. (2016). *Data literacy for educators.* New York: Teachers College Press.

17. Schwartz, D. L.; Chase, C. C.; Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759-775. http://dx.doi.org/10.1037/a0025140

18. Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London and New York: Routledge; Means, B., & Peters, V. (2014). *Lessons from funding five years of digital courseware.* Menlo Park, CA: SRI Education.

19. Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757-798. DOI: 10.1111/j.1551-6709.2012.01245.x

20. Dunlovsky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58; Schwartz, D. L.; Chase, C. C.; Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759-775. http://dx.doi.org/10.1037/a0025140

21. Dunlovsky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58; Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L., DeSelms, J., Seethaler, P. M., Wilson, J., Craddock, C. F.; Bryant, J.D., Luther, K., & Changas, P. (2013). Effects of first grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology, 105*(1), 58-77. http://dx.doi.org/10.1037/a0030127; Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science, 342*(6161), 935-937.

22. Kali, Y., & Linn, M. (2010). Curriculum design as subject matter: Science. In B. McGaw, E. Baker, & P. Peterson (eds.) *International encyclopedia of education (3rd edition),* Oxford: Elsevier. pp. 468-474; ; Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science, 342*(6161), 935-937.

23. Bransford, J., Brown, A., & Cocking, R. (1999.) *How people learn: Brain, mind, experience, and school.* Washington, DC: National Research Council; Dunlovsky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58; Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science, 342*(6161), 935-937; Kali, Y., & Linn, M. (2010). Curriculum design as subject matter: Science. In B. McGaw, E. Baker, & P. Peterson (eds.) *International encyclopedia of education (3rd edition).* Oxford: Elsevier, pp. 468-474; Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when and how.* London and New York: Routledge.

24. U.S. Department of Education, Office of Educational Technology. (2014). *Future ready schools: Building technology infrastructure for learning.* Washington, DC: Author.

25. Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. W., & Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist, 35*(3), 149-164; Fishman, B., Marx, R. W., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *Journal of the Learning Sciences, 13*(1), 43-76.

26. Helfand, D. (2005, February 7). Reading program didn't boost skills. *Los Angeles Times.* Retrieved from www.latimes.com.

27. Hansen, E. E., Llosa, L. L., & Slayton, J. (2004). *Evaluation of the Waterford Early Reading Program as a supplementary program in the Los Angeles Unified School District: 2002-03.* Planning, Assessment and Research Division Publication, (177).

28. Llosa, L., & Slayton, J. (2009). Using program evaluation to inform and improve the education of young English language learners in US schools. *Language Teaching Research, 13*(1), 35-54.

29. Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better.* Cambridge, MA: Harvard Education Press.

30. Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review, 29*(2), 255-270.

31. Silva, E., & White, T. (2013). *Pathways to Improvement: Using psychological strategies to help college students master developmental math.* Stanford, CA: Carnegie foundation for the advancement of teaching.

32. Institute of Education Sciences and National Science Foundation. (2013). *Common guidelines for education research and development.* Washington, DC: Author; Shadish, W.R., Cook, T.D., & Campbell, D.T. (2001). *Experimental and quasi-experimental designs for generalized causal inferences.* Berkeley, CA: Houghton Mifflin; What Works Clearinghouse. (2014). *WWC procedures and standards handbook.* Version 3.0. Washington, DC: Author.

33. What Works Clearinghouse. (2014). *WWC procedures and standards handbook. Version 3.0.* Washington, D.C.: Author.

34. Griffiths, R., Chingos, M., Mulhern, C., & Spies, R. (2014). *Interactive online learning on campus: Testing MOOCs and other platforms in hybrid formats in the University System of Maryland.* New York: Ithaka S+R.

35. Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better.* Cambridge, MA: Harvard Education Press.

36. U.S. Department of Education, Office of Educational Technology. (2013). *Expanding evidence approaches for learning in a digital age.* Washington, DC: Author.

Pearson
80 Strand
London
WC2R 0RL

Join the conversation
@Pearson

pearson.com