Center for
Technology Innovation
at BROOKINGS

March 2013

Education Technology Success Stories

Darrell M. West and Joshua Bleiberg

INTRODUCTION

Darrell M. West is vice president and director of Governance Studies and founding director of the Center for Technology Innovation at Brookings. His studies include technology policy, electronic government, and mass media.

Joshua Bleiberg is a research assistant in Governance Studies at the Brookings Institution and has a Masters Degree in Education Policy and Management from the Harvard Graduate School of Education.

In 2011, a student logged onto the online classroom Udacity to take the final exam for her introductory Physics class. Khadijah Niazi had overcome several barriers to finish that exam. She lived in Pakistan, which recently blocked access to YouTube, the site Udacity used to host its video lessons. Undeterred, she posted a plea for help on an Udacity message board saying "I am very angry, but I will not quit." Hours later several classmates from Malaysia, Portugal, and England attempted to find a workaround that would allow her to finish the class. Soon a Portuguese professor found a way to download the videos from YouTube and then upload them to a photo-sharing website that Kadijah could access. The next day she took the final exam. Even more amazing than the technology know-how is the fact that Kadijah was 11 years old and aced the college level physics class with the highest distinction. [1]

Advances in technology are enabling dramatic changes in education content, delivery, and accessibility. Throughout history, new technologies have facilitated the exponential growth of human knowledge. In the early twentieth century, the focus was on the use of radios in education. [2] But since then, innovators have seen technology as a way to improve communication, learning, and the mastery of instructional material.

The next generation of education technologies is facilitating substantial change. Education technologies are evolving beyond lecture and group work to games, simulations, and augmented reality. [3] Software is creating environments where students can direct the creation of their own knowledge with nearly invisible prompts from teachers.

One possible virtue of digital technology is the cost savings. During the Great Recession, the education service industry lost over one million jobs. [4] State and local governments cut education spending, and this had ripple effects throughout the sector. Today educators from universities to elementary schools face an even more difficult task than before with fewer available resources. Given the political climate of budget cutting, the likelihood of a restoration of funding to pre-recession levels in the near future is low. In this situation, educational technologies take on increased importance as they seek to help over-burdened teachers deploy the next generation of assistive technologies.

Education faces unique resource problems beyond financial issues. The school day has a finite length and instructional time is a precious commodity. American students spend less time in the classroom than many other countries elsewhere in the world. [5] Teaching is a complex job that includes a number of rote but time consuming tasks. Tools that facilitate the memorization of basic facts free up teachers to help students who need personalized interventions. Every extra minute spent teaching makes a difference over the course of the school year.

Assessment technology has advanced very little if at all since the invention of the optical scan answer sheet a half-century ago.

Recent advances in assessment technology have the potential to help teachers and students. Without feedback on performance, teachers can't know if students have grasped the lesson and policymakers won't know whether their reforms work. Assessment technology has advanced very little if at all since the invention of the optical scan answer sheet a half-century ago. New assessment technologies can help cut the costs of testing while others allow for reliable assessment in real time. Advances in testing can assess students in a low stakes environment.

While there are many innovations in education technology, this report highlights five education technology success stories. Each has demonstrated the ability to improve efficiency and effectiveness in education systems. From language teaching robots to educational games, each has the potential to help students and teachers. We review these education success stories in order to offer lessons on how education stakeholders can better serve students and add value to their learning.

Robot Assisted Language Learning (RALL)

Language instruction presents classic education policy challenges. It is resource intensive because it requires specialized materials and teachers with language mastery. Traditionally a scarcity of skilled second language teachers has constrained education institutions. But new technologies have the potential to change how students learn new languages. Much of the work in learning a new language requires repetition and memorization. The grammar and vocabulary of language provide a defined structure. These conditions allow robots to provide excellent support in secondary language acquisition.

Researchers from the Center for Intelligent Robotics (CIR) at the Korean Institute of Science and Technology (KIST) and the Pohang University of Science and Technology (POSTECH) developed

two robots to serve as English teaching assistants. They created two different models named MERO and Engkey. Engkey is short for English Disc Jockey and looks like a stout penguin. [6] Some Engkey models have an expressive face designed to simulate different emotions. Others have a monitor that allows teachers to teleconference into the classroom.

In South Korea as in many developing nations, it is difficult to attract qualified English teachers to remote islands or rural areas. The teleconferencing technology opens up geographically isolated classrooms. Engkey has "stereo vision" and the ability to move around the classroom and interact with students. [7] MERO meanwhile is a "head only" robot. It looks closer to popular depictions of a robot with large eyes and colorful exaggerated features. The head is able to rotate on a plastic base. The robots look friendly and non-threatening to children. In 2010 Engkey cost about $8,700, but the Korean government hopes to bring that price down as production increases. [8] The Korean Education Ministry would like all 8,400 kindergartens in the country to have an English language robot by the end of 2013. [9]

Engkey-Penguin Head Version
Source: http://www.cnn.com/2010/TECH/innovation/10/22/south.korea.robot.teachers/index.html; Center for Intelligent Robotics; Korean Institute of Science and Technology

MERO and Engkey work through advanced speech recognition software. The robots use transcribed speech of Korean children and the audio files of the *Wall Street Journal*. The system includes sounds Koreans are most likely to confuse for the correct phoneme. To account for unexpected errors, the system can statistically infer what the speaker intended to say based on the nature of the mistake and the context of the sentence. The system generates a number of hypotheses about the speaker's intended word choice and then ranks them. [10] The ranking process allows for the robots to understand and correct speech from young students learning a new language.

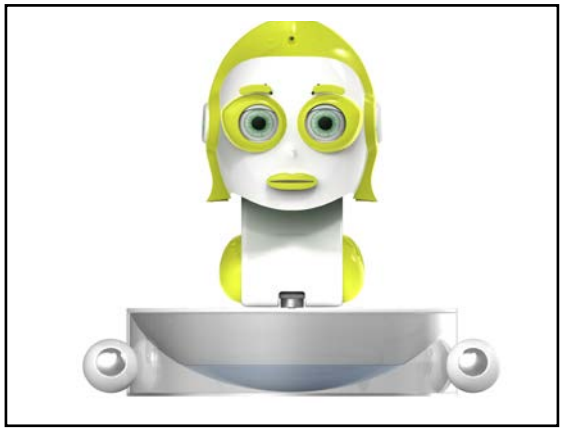
In South Korea as in many developing nations, it is difficult to attract qualified English teachers to remote islands or rural areas.

One of the greatest strengths of the language recognition system is how it identifies and corrects speaker errors. Early attempts at speech recognition software struggled because of unpredictable speaker errors that confused computer systems. Each robot includes a number of error rules to help account for likely mistakes. For example the system has in its database all of the English consonant and vowel noises (phonemes).

MERO and Engkey use a dialogue management system called RavenClaw. The RavenClaw code

includes conversation trees that allow the robots to create hierarchal maps. The electronic processor has protocols for pauses and turn-taking that make conversations sound real. The system includes a number of scripted conversations that students use to practice. RavenClaw also allows Engkey to correct the grammar, pronunciation, and vocabulary of English language learners. The system can identify the specific type of error the speaker makes and provide corrective feedback. RavenClaw can access an Example Expression Database and select a phrase closest to what the speaker intends and then correct the speaker. The system lauds students if the speaker uses proper language and engages in authentic conversations. [11]

Along with speech protocols, MERO and Engkey use facial expressions to communicate with students. They have a number of distinct expressive faces including like, dislike, neutrality, hope, fear, joy, distress, surprise, shame, and sadness. Engkey can also articulate gestures coordinated with facial expressions to wink, yawn, cheer, and sulk. The expressions make conversations more authentic. When the robots correct or praise a student, they also use the socially correct facial expression. Expressive emotions are an important component of verbal communication providing critical context to conversations. [12]



Mero
Source: http://irobotics.re.kr/pds/img/blog/mero_06.png; Center for Intelligent Robotics; Korean Institute of Science and Technology

The researchers who developed MERO and Engkey studied the effectiveness of robot-assisted language learning (RALL). The study included 21 students in a South Korean Elementary school who ranged from grade two to six. They were tracked into beginning and intermediate groups based on pre-test scores. All of the students were South Korean and spoke Korean as their first language. None of the students had lived in an English speaking country for longer than three months. The researchers developed lesson plans for both the beginning and intermediate classes. The unit centered on conversations appropriate to shopping in grocery and stationary stores. [13]

To measure cognitive effects of RALL on listening and speaking skills, students took a post-test in addition to the pre-test. To assess listening skill, there was a multiple choice test with 15 items. To demonstrate growth in speaking skills they conducted one-on-one interviews. The protocols included ten items to asses speaking skills. The rubric had a five-point scale for pronunciation, vocabulary, grammar, and communicative ability. [14]

The study suggests that RALL leads to large improvements in student speaking, but not listening skills. Student's scores on the post test were statistically significantly better than the pre-test. The effect size of RALL on speaking skills was large ranging from 0.86 to 0.9 standard deviations. If the average student who benefitted from the intervention started out with average skills, the results suggest they would finish well above average with better scores than eight out of ten typical students. The gains were persistent across the sub-categories of speaking skills: measured pronunciation,

vocabulary, grammar, and communicative ability. [15]

However, the study showed no statistically significant differences for listening skills. The researchers who conducted the assessment thought several factors may have influenced this outcome. One possibility is the text-to-speech components were not sufficient to mimic speech. It is conceivable that students who participated in scripted conversations were reading the script rather than responding to the robot. It is also possible that the various non-human sound effects the robots made had some negative effect on comprehension. [16]

Students reported a positive outlook on class time with MERO and Engkey. To evaluate student opinions, the KIST researchers designed a survey that students took both before and after the intervention. The survey included 52 items and responses were given on a four-point Likert scale, ranging from "strongly disagree" to "strongly agree". After completing the program students were significantly more likely to agree with the statement "You are interested in English." Student confidence also increased over the course of the program. They were more likely to report that "You can greet foreigners with confidence" and "You think that you can speak English better if you study harder." Students also reported increased motivation for learning English. They were more likely to agree that "You want to learn English more" and "You spend more time on studying English by yourself." Overall the students enjoyed the time they spent with MERO and Engkey. [17]

Secondary Language Acquisition (SLA) theory explains why RALL is successful at helping improve speaking skills. SLA theory proposes four different competencies necessary to improve conversation skills: comprehensible input, comprehensible output, corrective feedback, and motivation. The RALL lessons placed new words in a familiar context, which facilitates learning. The corrective feedback and praise that both MERO and Engkey provide helps student develop strong speaking skills. RALL encourages students to develop new constructs with context-supported lessons. [18]

In short, RALL is a powerful assistive technology with demonstrated impact. A scarcity of qualified secondary language teachers will likely persist into the future, and teachers will need support to instruct the next generation of students. Robots provide valuable opportunities for students to engage in authentic conversations. The success of MERO and Engkey in South Korea demonstrates how robots can have a positive impact on student learning.

Massive Open Online Course (MOOC)

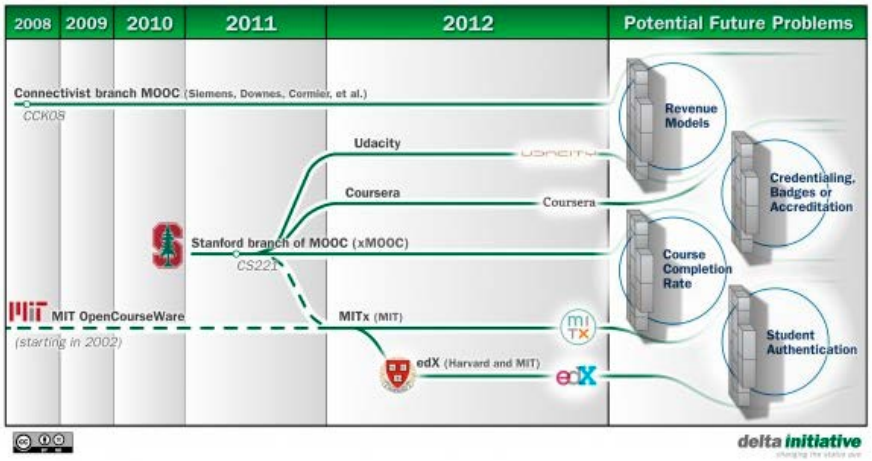
George Siemens and Stephen Downes coined the term "Massive Open Online Course" in 2008. In that year they facilitated arguably the first MOOC. They offered a free class called "Connectivism and Connective Knowledge" in partnership with the Extended Education and Learning Technologies Centre at the University of Manitoba. The class included over 2,300 students and used an RSS-aggregator as the discussion platform. [19] The term MOOC has come to define two substantively different online platforms. Connectivist MOOCs differentiate themselves with their approach to teaching. The MOOCs discussed in this paper define themselves through their affiliation with universities, financial backers,

and potential to generate revenue.

MOOCs have the potential to disrupt higher education, improving outcomes for students and expanding learning opportunities. Tuition has risen steeply over the past few decades, and the resulting cuts have negatively impacted students and restricted access by poorer students. MOOCs could dramatically decrease costs for universities and offer courses to students all over the world. [20]

MOOCs [Massive Open Online Courses] have the potential to disrupt higher education, improving outcomes for students and expanding learning opportunities.

The most prominent MOOCs are Udacity, Coursera, and Edx. MOOCs recently took a big step forward when the American Council on Education (ACE) recommended 5 Coursera classes for accreditation. ACE membership consists of more than 1,800 universities. [21] The recommendation means that accredited universities may soon accept Coursera credits towards completion of a degree. One possible scenario is that Coursera would charge a fee to verify the identity of a student and proctor a final exam using a web camera. ACE is currently in the process of a more comprehensive assessment of MOOCs but the recommendation signals the enormity of the changes that are coming. [22]



MOOC Growth
Source: http://www.deltainitiative.com/bloggers/is-higher-education-ready-for-rapid-evolution-of-xmoocs/attachment/fig-3-evolutioncombine20120927

Several universities already offer credit to students who complete MOOCs. The first university to offer MOOC classes was the University of Washington (UW). UW offers classes in applied mathematics, computer science, computational finance, and information security and risk management through the Coursera platform. [23] UW had previously offered certification programs through online classes and is now in the process of reformatting them for Coursera. [24] San Jose State University (SJSU) also has partnered with Udacity to offer a unique solution to the remediation of college freshman. Half of all freshmen can't pass introductory classes in core subjects at SJSU. This creates bottle necks in entry level classes and can contribute to a delayed graduation for students. Through Udacity, SJSU offers entry-level, math, college algebra, and elementary statistics for credit. The Udacity courses cost only $150, which is much less than a traditional credit. Udacity breaks its courses down into short video and quizzes which allows students to take the classes on their own timeline. [25]

The University of Wisconsin-La Crosse developed a math MOOC. Using a $50,000 grant from the Bill and Melinda Gates Foundation, UW La Crosse developed the open math course. [26] UW La Crosse wanted the course to serve high school students, those within the University system that wanted remediation, individuals preparing to re-enter a university, and those preparing to take a major gateway exam. The class roughly follows the same curriculum as the on-campus class MTH 051 Intermediate Algebra. [27] Many other universities have plans to offer credit for MOOCs.

MOOCs share three main characteristics. The first is nearly limitless scalability. Coursera had hundreds of thousands sign up for a single class. [28] Another attribute of MOOCs is their openness. Anyone can take a class because there is no admissions process. In addition MOOCs do not have tuition fees. In the future MOOCs will have to generate revenue. Possible options include charging tuition, certification exam fees, sponsorships for an individual's courses, or employee recruitment. A final defining feature is the technology that MOOCs use. In most cases they take advantage of open source or free platforms like Wikimedia or Google applications. The size of MOOCs allows for a model where each student pays very little rather than the traditional model where a relatively small group of students pays a substantial sum in tuition. [29]

The greatest strength of MOOCs over the traditional university model is the potential to reduce costs. Universities face a number of cost cutting challenges. Baumol's Cost Disease makes improving productivity difficult. Economic theory suggests that wages rise along with corresponding increase in worker productivity. But, in fields like education productivity increases do not have a discernible outcome for students. A mediocre professor can teach roughly the same number of students as a good professor and both of those professors will have some excellent students. Academic buildings with large lecture halls are expensive to maintain. MOOCs rely on inexpensive open source technologies and can scale very easily. In an online environment the costs for a one million-student class are not far off from a one hundred-student class. In a university setting MOOCs could decrease the number of sections for popular introductory courses, which would free up professors to teach other classes or complete research. [30]

Udacity, Coursera, and Edx incorporate retrieval exercises and deliberate practice into the curriculum. MOOCs also provide opportunities for constructivist learning and effectively use formative assessments. In a presentation on Coursera the company's founders Andrew Ng and Daphne Koller cite research that forms the base of the organization's pedagogical strategy. Jeffrey Karpicke and Janell Blunt found that "retrieval exercises" or active, cue-driven processes of reconstructing knowledge can serve as powerful teaching tools. The authors studied whether retrieval or "elaborate learning" (i.e. traditional studying techniques) had a greater impact on student proficiency. Students were taught the same lesson using different techniques and then assessed a week later. They found, "the advantage of retrieval practice over elaborative studying with concept mapping represent about a fifty percent improvement in long-term retention scores". [31] In another experiment students studied two different lessons. Students who studied using retrieval techniques scored significantly higher on a short-answer test. [32]

Ng and Koller acknowledge the work of Louis Deslauriers, Ellen Schelew and Carl Wieman. The authors conducted an experiment with two different sections of an introductory physics class. One classroom used traditional teaching methods and one using the principles of "deliberate practice." Deliberate practice is a concept that incorporates constructivism and formative assessment. They found that students in the experimental section were more likely to engage and were more likely to attend class. [33]

MOOCs open up learning opportunities for all people. The value of providing a classroom for students to learn is incalculable. MOOCs provide a platform for professors to communicate their theories to a large audience. MOOCs allow world class professors to teach their research with people all over the world. The opportunity to communicate with a nearly unlimited pool of people will accelerate the transfer of knowledge.

As the cost of university attendance increases, MOOCs will draw greater interest. Higher education is entering a new era where institutions will have unparalleled reach. New classroom technologies have the capability to democratize knowledge and expand learning opportunities. Udacity, Coursera, and Edx will offer the next generation of students a cornucopia of knowledge.

Minecraft

The Internet revolutionized how teachers and students share information. It placed teaching standards, lesson plans, and supplementary materials at the fingertips of teachers. Despite the enormity of this change the technology of instruction has remained static. The Internet has facilitated but not changed the standard teaching techniques: lecture, group work, individual reading, and slide presentation. New technologies have created new media for teachers to instruct students. The computer game Minecraft is the first step in that direction. Minecraft is a dynamic game that has nearly inexhaustible flexibility.

Minecraft is a highly successful "sandbox" computer game. It is open-ended with no defined narrative or game play objectives. Players approach the game similar to the way children play with Legos or blocks. The game play of Minecraft is deceptively simple. On its face players gather resources and build things with those resources. The worlds that Minecraft generates for players are unique and very large. The cubic volume of a Minecraft world is two hundred sixty-two quadrillion, one hundred and forty-four trillion "blocks." If each block had sides one meter long then a Minecraft world would have a greater surface area than Neptune. [34]

Minecraft is a highly successful "sandbox" computer game. It is open ended with no defined narrative or game play objectives.

The computer-generated worlds of Minecraft reflect many of the same features as Earth. Minecraft has over a dozen biomes including forest, desert, plains, swamp, jungle, ice plains, taiga, mountain, and ocean. Each biome has climate specific weather and the entire planet has a normal day/night cycle. The world properly simulates gravity and watersheds. Everything in the world consists of blocks, which include different types of dirt, wood,

minerals, precious metals, water, and many others. Physics and chemistry adhere to basic principles. Players can gather natural resources to build nearly anything their imagination can conjure. With the proper resources it's possible to build functional electrical circuits or coal-powered engines. The game creates a reasonable facsimile of Earth with its bountiful resources and laws of nature. [35]

Minecraft has several features that allow it to serve as a teaching tool. The game has a modification or mod called MinecraftEdu. MinecraftEdu gives teachers the tools to conduct a lesson in a virtual world. The mod allows teachers to place text around the world and control the location of students. [36] A community of teachers and educators has grown to support teaching on the Minecraft platform. They have created a dedicated Minecraft teaching wiki including lesson plans and research. [37] As a teaching platform Minecraft has tremendous potential. In social studies, students could build their own pyramid and during the process encounter the same engineering challenges as the Egyptians. Students could recreate scenes from literature. Teachers could demo the principles of biology, physics, and chemistry in real time. Minecraft enables students to engage in constructivist learning. According to constructivist theory, students that engage their environment gain trustworthy information through their own exploration. The Minecraft simulations provide an intellectual playground that encourages authentic learning. MinecraftEdu provides a brand new instructional media.

Jeffrey Adams is a middle school science teacher at the Crescent School in Toronto. Adams used Minecraft to teach a lesson on sustainable planning. Students were asked to create a model of a sustainable city that used little energy. Many of the students chose to build their models virtually in Minecraft. Students were able to build their cities as they saw fit and were given a particular role of farmer, builder, or miner. Adams wanted to find if using Minecraft would increase engagement and performance. Towards that end he administered a survey before and after the unit. Eighty-six percent of students reported they wanted to use the game again in school. Eighty percent believed that gaming allowed them to be more creative and do things they couldn't have done otherwise. Overall the students found the Minecraft format liberating. [38]

Students use MinecraftEDU
Source: http://minecraftedu.com/#!prettyPhoto

Karen Yager and Andrew Weeding are teachers at the Know Grammar School in Sydney, Australia. They used Minecraft in her classroom to teach 12-13 year old boys in a science class. Students were asked to create a model of a city that used little energy. Students were free to choose their presentation format and nearly all (ninety-six percent) chose Minecraft. They had five days to build their virtual cities and had daily challenges in English, mathematics, science and foreign language to ensure the project was multidisciplinary and to keep the students focused. Students then presented their virtual cities at the end of the unit. They found that the boys enjoyed the freedom to explore and take risks within a structured environment. [39]

Professor Jessica Bayliss used Minecraft as a teaching platform in her class Game Artificial

Intelligence at the Rochester Institute of Technology. Students in the class had to create new terrains and non-player characters in addition to other tasks. College and graduate students were able customize the game extensively. The source code for the game is in Java which is a common programming language. In addition, the extensive modding community provided creativity and support. Bayliss had taught the course using other software platforms in the past but found that students responded better to Minecraft. [40] Bayliss found that using Minecraft increased student motivation for artificial intelligence through modding. [41] She notes that a higher number of students asked to complete AI-related independent studies than in past years. Students were able to engage the Minecraft platform and their own creativity.

Map of Roman Coliseum for Minecraft Lesson
Source: http://warrenbez.com/wp-content/uploads/2012/03/col2.png

Minecraft succeeds as a game for the same reason that it succeeds as an instructional tool. It builds upon a context familiar to everyone or the rules of nature. Keeping the game simple allows for players to engage in authentic learning. The constructivist base allows individuals of varying skill levels including elementary school and graduate students to learn from the game. This dynamic classroom will provide teachers a new instructional technology. As MinecraftEdu matures the opportunities for learning will only increase.

Computerized Adaptive Testing (CAT)

The technology of standardized testing has changed very little over the past fifty years. Psychometricians have made some advances in terms of item writing and test design. However, the paper-based test (PBT) with an optical answer sheet has endured. Computerized Adaptive Testing has the potential to revolutionize the stagnant field due to its greater precision and practical advantages over traditional test formats.

PBTs have numerous weaknesses. The ability to complete a PBT may require skills not valid to an inference about subject proficiency. For example a student with dyslexia may have very strong math skills but struggle to read test items. A student with poor fine motor control may have difficulty filling in test bubbles but may have mastered the subject. The process of grading PBTs is prone to error and expensive. The printing, storing, delivery and pickup of millions of paper tests contribute to the cost of standardized testing. Grading handwritten essays requires a huge logistical effort. Furthermore, the PBT format creates many opportunities for cheating. The true extent of cheating on standardized tests is an understudied phenomenon. But, it is reasonable to assume that as more education stakeholders find their personal fortunes tied to test scores more will cheat. Securing tests requires a huge commitment of school resources. These issues are inherent to the PBT format and difficult to solve.

CAT combines two old technologies: computers and adaptive algorithms. The first test to adapt test items to the test taker was the Stanford-Binet IQ Test in 1905. [42] Modern CAT tests combine both of these approaches. CAT selects test items based on how the test taker answered earlier questions. Many different education organizations successfully use CAT. Today graduate students can take the Graduate Records Examination (GRE) and Graduate Management Admission Test (GMAT) in the CAT format. Currently Idaho, South Dakota, Oregon, and Delaware use CAT. Delaware is the only state to use CAT in an accountability setting. [43] CAT will also play a role in the new Common Core State Standards (CCSS) assessments. The Smarter Balanced Assessment Consortium has committed to using CAT although the other CCSS consortia, Partnership for Assessment of Readiness for College and Careers (PARCC), have decided to use fixed form tests. [44]

CAT has more reliable test scores than PBT. PBTs have the same set of questions for each student. For the majority of students (those who are average) PBTs allow for a decent snapshot of performance. However, for students at the tail ends of the distribution (very strong and very weak students) there is evidence of decreased reliability. CAT asks normatively better questions than PBTs and adapts the test items to the student's skill level. If the test taker answers a question correctly then the next question is more difficult and if he or she answers an item incorrectly then the next question is easier. Test scores have greater reliability when the difficulty of the item closely mirrors the knowledge of the student. To best assess students, tests need "Goldilocks" items; items that are not too hard, or too easy, but just at the student's skill level. Imagine an average 8th grade math student who then takes a 7th grade math test and 9th grade math test. The student will likely do very well on the 7th grade test and very poorly on the 9th grade test. Unfortunately neither of those scores allows for a strong inference about the students proficiency in 8th grade math. Test items with an appropriate level of difficulty make for a better test.

Researchers have confirmed that CAT scores have greater reliability than PBT at the extreme ends of the distribution. Gage Kingsbury and Carl Hauser compared "test information and measures of score accuracy and classification between fixed and adaptive forms of 4th and 8th grade reading and mathematics tests and noted that information at the extremes of the ability distribution was three times greater in the adaptive setting." [45] In another study of CAT in Oregon, Tony Alpert found smaller standard errors for CAT than for PBT at the tail ends of the distribution. [46] The virtue of reliability is increased confidence that students would get a similar test score if they retook the test. If students ace a test and then fail a similar version then it has very low reliability and little diagnostic value. As policymakers create higher and higher stakes for test scores the merit of reliability only increases.

CAT costs significantly less to administer than PBT. Gary Phillips, Vice President and Chief Research Scientist at American Institutes for Research, in testimony before Congress estimated that CAT is half as expensive PBTs. Phillips argued that the costs of printing, transferring to schools, and grading inflated the price of PBTs. [47] The Brown Center on Education recently estimated that states spent roughly $1.7 billion on standardized testing each year. [48] Schools may have elevated costs in the short term associated with upgrading or purchasing computers. However over the long term schools

will have large savings.

CAT takes students less time to complete which frees up the school day for instructional time. Walter Way found that CAT needs fewer test items to achieve a similar level of precision to PBT. If questions are tailored to the test taker then time is not wasted or unnecessarily easy or difficult questions. Way also found that students can complete some adaptive tests forty to fifty percent faster than paper tests. [49] Given the dearth of instructional time in American schools spending fewer minutes on testing is an attractive option for policymakers.

Another advantage of CAT is the option to include confidence-boosting test items. Test taking causes anxiety in students. Standardized tests ought to have high stakes but stress can also impede academic performance. Because CAT tracks students throughout the test it can include un-scored confidence boosting items. None of the organizations or states use test items in this way but their inclusion would be relatively simple. Joachtm Hausler and Markus Sommer studied confidence building items and found students had greater self-confidence when motivator items were included. Furthermore they were unable to find evidence that the inclusion of these items biased test results. [50] The inclusion of confidence boosting items has the potential to help decrease test anxiety with few tradeoffs.

A final benefit of CAT is the ease of use for special education students. Many students with disabilities could demonstrate proficiency on paper based tests but require accommodations such as an aid to read the test questions. Computers can easily convert text to speech, speech to text, and magnify test items. Simple accommodations could allow many students to participate in the same accountability systems as general education students. This saves resources but more importantly treats all students with the same standard.

PBT served as the dominant test format for decades. Computers have now improved to the point where they are a far superior platform. They allow for more precise assessment for less money. CAT frees up valuable instructional time and can help limit student anxiety. More and more education stakeholders are using student test scores to drive policy decisions from school closures to tenure granting. As this trend continues education stakeholders make a tacit promise to administer reliable tests. Using CAT is part of fulfilling that promise.

Stealth Assessments

Assessing student growth is an essential component of high quality teaching. The two main types of assessments are formative and summative. Formative assessments typically occur during a lesson or unit. They evaluate a student's strengths and weaknesses so that the teacher can tailor their practice accordingly. Summative assessments are end of the unit tests that provide feedback to students. The difference is subtle and relates to the inference an educator wants to make or how they will use the data. Generally speaking, summative assessments are high stakes. For example if a student fails a graduation exam they will not receive a diploma. Formative assessments are sometimes though

not always low stakes. They offer an opportunity to course correct mid-journey without punitive actions. The data formative assessments provide is critical to high quality teaching.

Formative assessments share many of the same problems as standardized tests. Administering formative assessments represents another strain on instructional time. Designing assessments that are relevant for a particular class and not summative takes time and skill. Taking full advantage of formative assessments requires reflection on student needs and then editing future lessons. Formative assessments are powerful but highly resource-intensive teaching tools.

Stealth assessments represent a leap forward in assessment technology. Stealth assessments embed formative assessments into games. Valerie Shute, a Professor at Florida State, and her colleagues developed the first stealth assessment and intended them not to deceive students but rather capture data unobtrusively. [51] They collect data about student learning that teachers can then use to improve and individualize instruction. Shute has discovered many benefits of stealth assessment.

Stealth assessment addresses some classic education resource problems. From a psychometric perspective the low stakes environment adds value. Psychologists have long understood that studying how children play allows for insight into the learning processes. Stakes change behavior and students use different strategies to solve problems in a sandbox than they do when taking the SAT. Stealth assessments capture different data than high stakes assessment because the students' behavior changes when engaged in a game as opposed to when focused explicitly on an assessment. Another benefit of games is how they motivate students. Games engage the player in a narrative and the player becomes part of the storytelling. Playing games motivates students to push further even when encountering challenges. Conversely, paper and pencil tests do not motivate students. Students' willingness to play stealth assessment games out of the classroom could provide a huge boon to teachers because it would provide data without using class time.

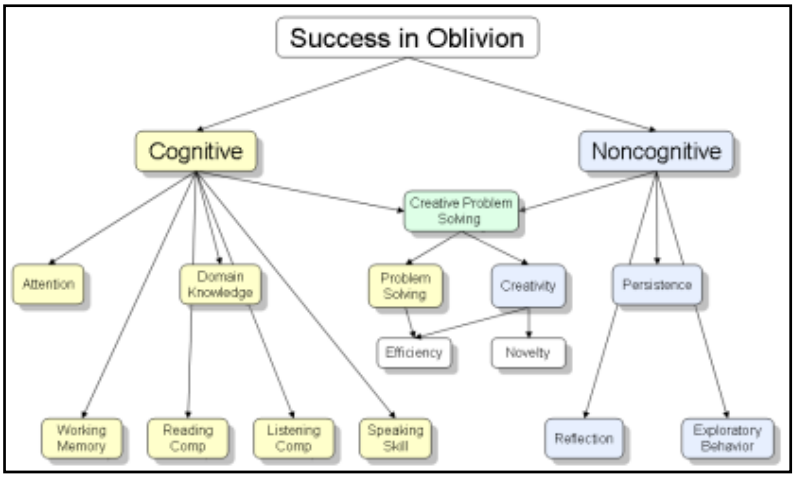


Illustration of a Competency Model for Success in the Game *Oblivion*
Source: Shute, Valerie J., et al. "Melding the power of serious games and embedded assessment to monitor and foster learning." Serious games: *Mechanisms and effects* (2009): 295-321.

Shute has developed as a proof of concept how to collect data from popular off-the-shelf video games such as "The Elder Scrolls IV: Oblivion." [52] "Oblivion" is a classic "Dungeons & Dragons" style open world role-playing game. Shute and her colleague built a Bayesian model to analyze player actions in the game for evidence of various competencies. They used the model to examine students' problem solving skills. In "Oblivion" players had to cross a river with dangerous fish. Players had several options including using magic, finding

a bridge, or swimming across. Using game data, their model can evaluate the novelty and efficiency of how players responded. Such data may appear irrelevant but in fact it allows educators a view of students' competencies. Shute and her team are now developing a new game called "Newton's Playground" to evaluate high school students' proficiency with physics. [53]

Another example of a stealth assessment is "Save Patch" developed by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). They designed the game to help 4th, 5th, and 6th graders learn how to use fractions. They had some success in achieving that goal. Students had higher math scores on a post-test when compared with the control group, although the differences were small. The "Save Patch" experiment also had two unexpected achievements. First, they were able to gather in-depth data on students' competency in solving fraction problems from the game itself. Second, they found that students with low self-efficacy in math scored significantly better after playing the game. "Save Patch" demonstrates how a game can generate data that enables targeted interventions for students. [54]

Current education reform efforts have focused on developing the infrastructure to gather large data sets on student learning. The next phase of education reform will center on how to leverage that data to improve instruction. These new data systems have tremendous potential to help revolutionize education. Incorporating data from stealth assessments will serve as a critical component to make these new systems work better. Ultra-low stakes assessments can provide a more holistic view of student competencies without significant demands on school resources.

Conclusion: Abandoning the Egg Crate Approach to Education

In Dan Lortie's book *Schoolteacher*, he describes the cellular structure of the American education system. In Colonial times teachers worked in one-room schoolhouses. They were isolated from each other and were wholly responsible for teaching all subjects and all ages. Most children did not receive a formal education and thus schools were spread out over a rural undeveloped America. Schools were "organized around teacher separation rather than teacher interdependence." [55] The nascent education system consisted of a series of independent cells. As the cities grew so did schools. Students were divided into age groups and assigned to different classrooms.

...[T]he egg crate culture remains the dominant norm in charter, public, and private schools to this day.

However, this growth of schools did not lead to dramatic changes in the teaching profession. Classrooms were stacked on top of each other like "egg crates." Teachers each occupied their own cell and rarely interacted with each other. The norm to contain teaching in individual classrooms has continued unabated. Recently the choice movement offers parents more options for their children. Many charter school operators value collaboration between teachers. Despite this shift, the egg crate culture remains the dominant norm in charter, public, and private schools to this day. [56]

The main force behind the cellular school is bureaucratic momentum. The egg crate model of schooling places unnecessarily high levels of pressure on

teachers. Assigning teachers a piece of turf upon which he or she is responsible for everything stifles collaboration and creativity. Policymakers ought to encourage teachers to gain expert knowledge in specific subjects and with instructional techniques. Teachers ought to serve as the leaders of flexible education communities where collaboration rather than isolation is the standard. New technologies can disrupt old norms and help develop new ones.

The technologies described here could loosen the demands on educators freeing them to teach. Imagine a school where teachers work in synchronization with each other to educate students. Teams of teachers each with their own special skills would work jointly to develop curricula and specialized interventions. A school where students could visit the cafeteria with robots to learn words for different foods, a reality simulation lab where students could work on different projects, and a library where students took micro-lessons through MOOCs to advance or reinforce learning. In such a school, students could direct much of their own learning. This report demonstrates these technologies are not science fiction.

Critics often fear that new technologies will cost education professionals their jobs. Technology can serve as both a creative and a destructive force. In some cases a new technology will change an industry fundamentally. In a world with cars there is no need for horse drawn carriages. There is reason to believe that technology will supplant rather than destroy the teaching profession. Educators have a uniquely complex job and have far more to do than time allows. New technologies will take over some teacher responsibilities. However, technology can't make a teacher obsolete in the same way as a carriage. In the imaginary school described above teachers still play an essential role in directing student learning. If anything, technology may enable teachers to focus on higher-level learning issues.

In a higher education setting, technology has numerous benefits. MOOCs could lead to a decrease in the number of introductory survey courses. This would free up professors to offer more advanced seminars and work on their own research. MOOCs present a new platform to present research. Professors could create MOOCs to share their recent discoveries instead of writing a paper. This would reach a considerably larger audience and allow professors to communicate with interested parties in the field.

Opponents of technology in the classroom often cite high purchasing and retraining costs. Many of the technologies discussed here have large upfront costs. Policymakers should consider the cost of new policies not merely in terms of the current fiscal year but also the long-term budget implications. States and districts should also consider engaging other organization to bulk order technologies and negotiate the price down. Production costs for new technologies will often drop considerably as the market develops.

We need policy changes that facilitate education innovation. In many public K-12 schools, rules designed for an agrarian or industrial world limit flexibility. School finance focuses on seat-time requirements which apportion money to schools based on the number of days students sit in a classroom. Grade promotion is based on time in class as opposed to skill mastery. We need flexible structures that emphasize learning as opposed to time spent in classrooms.

In addition, some K-12 schools and colleges discourage distance learning by not allowing credit for those classes or taking money away from districts when their students enroll in distance courses. These types of policies discourage innovation and make it difficult for students to access valuable learning materials through new delivery systems.

In countries all over the world students are learning English. Many of these schools lack qualified English teachers due to poverty or geographic isolation. Engkey can supplement language instruction but also helps connect teachers with students. New Engkey models include monitors that allow teachers to teleconference from anywhere in the world. With Engkey nearly every school has access to highly qualified language instructors.

Universities can use MOOCs to clear the enrollment logjams for entry-level survey courses. A little over half of students graduate from 4-year public universities in 6 years.[57] One factor contributing to this grim statistic is the unavailability of courses required for graduation. Students shouldn't have to pay an entire semester's tuition because a class is full. MOOCs provide flexibility to students and remove bureaucratic hurdles from students graduating on time.

Forty-five states have adopted the Common Core State Standards and will use the new assessments currently in development by PARCC and the Smarter Balanced consortia. CAT is valuable because it allows for more precise assessments of all students with few tradeoffs. Unfortunately PARCC has chosen to use a fixed form computer based test. It is possible PARCC has chosen a fixed form test because of the lower development costs associated with a smaller item bank. However, PARCC and the states not participating in the Common Core should change course and use adaptive tests. Innovations in standardized testing occur rarely and educators ought to take advantage of this policy opportunity.

Lawmakers need to update privacy rules to allow for data collection from games outside of the classroom. Districts should develop tools to store this data into existing education data warehouses. Educators must communicate with parents to earn the permission and capture data from the games students play to incorporate into assessment systems. These changes will maximize the value of stealth assessments.

Leaders should rely upon empirical research to guide budget decisions. Shrewd decision making with evaluation can bring down costs. A unique feature of the technologies discussed in this report is their ability to collect data. In the coming years education will see an explosion of research as "big data" becomes available. The main barrier to the creation of those data sets is data collection. Harnessing new technologies will allow for better data capture that can improve instruction and illuminate which programs work effectively. Eliminating inefficient programs can in turn save schools money. Policymakers should consider not only the price of a new technology but also its transformational potential.

Technology can create a truly meritocratic education system. In the status quo we assume students need the same amount of time in the classroom to learn the same material. If a student can demonstrate mastery of a college level physics class through self-directed learning then spending

hundreds of hours in a high school class is a questionable policy. New assessment technologies can more precisely detect student mastery. Furthermore the opportunities for evaluators to gather data from games and other assessments will only increase. Providing students more opportunities to learn and to demonstrate proficiency will allow for a more meritocratic education system. Assessments don't have expectations for students. Any student regardless of background can use technology to earn accreditation for their learning.

Endnotes

1. Ripley, Amanda. "College is dead. Long live College!" *Time* 18 Oct. 2012: n. page. Web. 15. February 2012.
2. Reid, Seerley, and Daniel Day. "Radio and records in education." Review of Educational Research 12.3 (1942): 305-322.
3. West, Darrell M. *Digital Schools: How Technology Can Transform Education*. Brookings Institution Press, 2012.
4. Haver Analytics. "Employment Household Survey" Employment: Education Services January 1983 to January 2013
5. *Expanded Learning Time by the Numbers*. Center for American Progress, April 2010. Web. < http://www americanprogress.org/issues/education/news/2010/04/22/7716/expanded-learning-time-by-the-numbers/>
6. Sang-Hun, Choe. "Teaching Machine Stick to Script in South Korea" *Time* 10 July 2010: n. page. Web. 15. February 2012.
7. Lee, Sungjin, et al. "Cognitive effects of robot-assisted language learning on oral skills." *INTERSPEECH 2010 Satellite Workshop on" Second Language Studies: Acquisition, Learning, Education and Technology*. 2010.
8. Horn, Leslie. "South Korean Schools Testing Robot Teachers." *PC Magazine* 28 Dec. 2010: n. page. Web. 15. February 2012.
9. Ibid Sang-Hun, Choe
10. Lee, Sungjin, et al. "Affective effects of speech-enabled robots for language learning." *Spoken Language Technology Workshop (SLT), 2010 IEEE*. IEEE, 2010. B
11. Lee, Sungjin B
12. Lee, Sungjin A
13. Lee, Sungjin A
14. Lee, Sungjin A
15. Lee, Sungjin A
16. Lee, Sungjin A
17. Lee, Sungjin B
18. Lee, Sungjin B
19. Buck, Tara. "The Massive Effect of MOOCs on Higher Education." *Ed Tech Magazine* 24 Jan. 2013. 15. February 2012. Web. <http://www.edtechmagazine.com/higher/article/2013/01/massive-affect-moocs-higher-education>
20. Martin, Fred G. "Will massive open online courses change how we teach?." *Communications of the ACM* 55.8 (2012): 26-28.
21. "About the American Council on Education." 2013.Web. <http://www.acenet.edu/about-ace/Pages/default.aspx>.
22. Kolowich, Steve. "American Council on Education Recommends 5 MOOCs for Credit." *The Chronicle of Higher Education* 7 February 2013. Web. <http://chronicle.com/article/American-Council-on-Education/137155/>.
23. "University of Washington Is First in U.S. to Provide Credit Classes and Certificate Programs on MOOC Platform." 18 July 2012. Web. <http://www.pce.uw.edu/newsroom.aspx?id=10865>.
24. Long, Katherine. "UW to offer fee-based courses through Coursera." *Seattle Times* 18 July 2012. Web. <http:// seattletimes.com/html/localnews/2018714077_coursera19m.html>.
25. Das, Sumi. "Udacity, San Jose State University offer online classes for credit." 29 January 2013. Web. <http://news. cnet.com/8301-1023_3-57566552-93/udacity-san-jose-state-university-offer-online-classes-for-credit/>.
26. "University of Wisconsin-La Crosse to develop math MOOC to boost college readiness and success." 13 November 2012. Web. <http://www.wisconsin.edu/news/2012/r121113.htm>.
27. "College Readiness: University of Wisconsin System's Math MOOC." 2013. Web. <http://www.uwlax.edu/mathmooc/ index.html>.
28. Adams, Susan. "Is Coursera the Beginning of the End for Traditional Higher Education?" *Forbes* 17 July 2012. Web. <http://www.forbes.com/sites/susanadams/2012/07/17/ is-coursera-the-beginning-of-the-end-for-traditional-higher-education/>.
29. Rodriguez, C. Osvaldo. "MOOCs and the AI-Stanford like Courses: Two Successful and Distinct Course Formats for Massive Open Online Courses." *Learning* (2012).
30. Ruth, Stephen. "Can MOOC's and Existing E-Learning Efficiency Paradigms Help Reduce College Costs?." *Available at SSRN 2086689* (2012).
31. Karpicke, Jeffrey D., and Janell R. Blunt. "Retrieval practice produces more learning than elaborative studying with concept mapping." *Science* 331.6018 (2011): 772-775.
32. Ibid Karpicke and Blunt

33. Deslauriers, Louis, Ellen Schelew, and Carl Wieman. "Improved learning in a large-enrollment physics class." *Science* 332.6031 (2011): 862-864.
34. Hafer, TJ. "Minecraft Infograhic reveals total volume of one entire world." *PC Gamer* 6 February 2013 Web. <http://www.pcgamer.com/2013/02/06/minecraft-infographic-reveals-total-volume-of-one-entire-world/>.
35. Short, Daniel. "Teaching scientific concepts using a virtual world—Minecraft." *Teaching Science-the Journal of the Australian Science Teachers Association* 58.3 (2012): 55.
36. "MinecraftEDu." 2012 Web. <http://minecraftedu.com/mod.php>.
37. "MinecraftEDu Wiki."2012 Web. <http://minecraftedu.com/wiki/index.php?title=Main_Page>.
38. Adam, Jeffrey "Teaching Boys at the Coal Face: Mining Keep Pedagogical Approaches." *Action Research Report* (2012): Page 5.
39. Yager, Karen and Weedings, Andrew. "Teaching Boys at the Coal Face: Mining Keep Pedagogical Approaches." *Action Research Report Approaches* (2012): Page 105.
40. Bayliss, Jessica D. "Teaching game AI through Minecraft mods." *Games Innovation Conference (IGIC), 2012 IEEE International*. IEEE, 2012.
41. Ibid Bayliss
42. "First Adaptive Test." 2012 Web. <http://iacat.org/node/442>.
43. Wang, Shudong, et al. "Construct Validity and Measurement Invariance of Computerized Adaptive Testing: Application to Measures of Academic Progress (MAP) Using Confirmatory Factor Analysis." *Online Submission* (2012). Web. <www.doe.k12.de.us/dcas/files/Vol4_RelVal_Report.pdf>.
44. "Computer Adaptive Testing." 2013 Web. <http://www.smarterbalanced.org/smarter-balanced-assessments/computer-adaptive-testing/>.
45. Quoting Stone, Elizabeth, and Tim Davey. "Computer-Adaptive Testing for Students With Disabilities: A Review of the Literature." (2011). Page 6.
Stone Davey cite Kingsbury, G. Gage, and Carl Hauser. "Computerized adaptive testing and No Child Left Behind." *Annual Meeting of the American Educational Research Association, San Diego, CA*. 2004.
46. Alpert, Tony. "A coherent approach to adaptive assessment." *National Academies, Board on Testing and Assessment and The National Academy of Education Workshop, Best Practices for State Assessment Systems, Washington, DC. Retrieved from http://www7. nationalacademies. org/bota/Best_Practices_Workshop_2_Agenda. html*. 2010.
47. Phillips, Gary W. "Testimony for the US Senate Committee on Health, Education, Labor and Pensions." (2010).
48. Chingos, Matthew M. "STRENGTH IN NUMBERS." (2012).
49. Way, Walter D. (2010, June). *Some perspectives on CAT for K-12 assessments*. Presented at the National Conference on Student Assessment, Detroit, MI.
50. Hausler, Joachtm, and Markus Sommer. "The effect of success probability on test economy and self-confidence in computerized adaptive tests." *Psychology Science* 50.1 (2008): 75.
51. Shute, Valerie J. & Ventura, M. (in press). Measuring and supporting learning in games: Stealth assessment. White paper for *MIT series*, published by the MacArthur Foundation. Web. <http://myweb.fsu.edu/vshute/pdf/white.pdf>.
52. Shute, Valerie J., et al. "Melding the power of serious games and embedded assessment to monitor and foster learning." *Serious games: Mechanisms and effects* (2009): 295-321.
53. Ibid Shute Ventura 2013
54. Delacruz, G. C. (2011). *Games as formative assessment environments: Examining the impact of explanations of scoring and incentives on math learning, game performance, and help seeking*. (CRESST Report 796). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
55. Lortie, Dan C. *Schoolteacher: A sociological study*. Chicago: University of Chicago Press, 1975. Page 14
56. Ibid
57. NCES IPEDS 2012 Web. <http://nces.ed.gov/ipeds/>.

Governance Studies
The Brookings Institution
1775 Massachusetts Ave., NW
Washington, DC 20036
Tel: 202.797.6090
Fax: 202.797.6144
www.brookings.edu/governance.aspx

Editor
Christine Jacobs
Beth Stone

Production & Layout
Beth Stone

Email your comments to gscomments@brookings.edu