Essay 11

New Frontiers in Technology for Assessment in Low-Income Contexts



Chief Technical Officer, International Development Group, RTI International

Carmen Strigel

Director, ICT for Education and Training, RTI International



"We need assessment approaches that inform and guide children's learning progress, and stay current with the skills and content being taught."¹

In education circles, consensus is growing that commonly used assessments aren't capturing the breadth of what children are (or should be) learning. A recent Brookings Institution study indicates that, while educators and education stakeholders worldwide share many of the same expectations of what skills are important for future generations, countries are at very different stages and are using different approaches to get there.²

Assessing students' skills—as one moves away from basic reading, writing, and mathematics towards critical thinking and problem-solving skills—becomes increasingly difficult with standard pencil and paper assessments. Assessing such complex skills using traditional methods also strikes experts as less and less authentic. However, it is costly and training-intensive to try to assess a group of students' problem-solving or team-working skills via close observation. We claim that the use of technology for assessment has great potential, especially for assessing complex skills, and including in low-income contexts.

TECHNICAL CHANGES BROADEN OUR POSSIBILITIES

To date, the use of technology for assessment has focused on improving data quality and the efficiency of assessment. Electronic testing has a long history in large-scale standardized student assessment in high-income countries, for example via the cognitive Scholastic Aptitude Test (SAT) or Graduate Record Examinations (GRE). Likewise, electronic platforms to facilitate oral early reading and numeracy assessments are widely used for sector diagnostic and program evaluations in low-income settings. Such technology facilitates efficient data collection and management and allows experts to make assessment data actionable through just-in-time analysis and built-in decision support for non-technical decisionmakers.

Advances in the computational power of handheld devices and psychometric tools have dramatically facilitated adaptive testing. Although such methods have been used in large-scale standardized tests (e.g., the Graduate Management Assessment Test, or GMAT), they can also be used for self-assessment and self-directed learning. Since the early days of drill-and-practice software programs, computer-assisted learning has proven its utility by virtue of being responsive to certain objective forms of input, thus allowing individuals to gauge their own progress and skills gaps. The virtue of adaptive testing is that, at least in principle, it can offer the promise of instruments with validity and reliability at both ends of a cognitive scale, and have practical utility for countries as diverse as Finland or Burundi.

LEVERAGING TECHNOLOGY TO BETTER ASSESS "COMPLEX" SKILLS

We argue that technology can bring assessments closer to revealing the true manifestation of the skills or abilities under investigation. Technology also offers opportunities for new assessment methodologies that provide a more differentiated and holistic understanding of a child's capabilities and needs.

By leveraging technology to better approximate real world situations where assessed abilities are manifest, we hope to address two inter-related problems in traditional assessments: Measurement impurity and ecological validity.

Measurement impurity refers to the phenomenon whereby scores that result from an assessment represent individual differences in both the primary construct of interest, as well as a host of correlated processes or constructs. For example, measures of inhibitory control index individual differences in inhibitory control, as well as individual differences in processing speed and receptive language (which informs the ability of a child to understand the task). Measurement impurity thus undermines the very premise of assessment as we may not be measuring what was intended.

Technology may help detect the behavioral and/or physiological signa-



tures that accompany many cognitive constructs. The ability to acquire and rapidly integrate these sources of information with task performance allows for more precise measurement and can potentially refine the characterization of the construct under study. This gets us closer to answering fundamental questions about whether a student's progress (or lack thereof) may be hindered by underlying difficulties or under-developed non-cognitive processes. An example of technology-mediated multi-modal assessment is the work done by researchers at Stanford University combining data on student gestures from a Kinect sensor with data logged from a tangible user interface, while students worked in pairs to solve a cognitive task.

The ecological validity problem refers to the disconnect between the contexts in which cognitive assessments often occur (i.e. emotionally neutral, one-on-one, minimal distractions) and the contexts in which the assessed abilities and skills are drawn upon (i.e., emotionally variable, social settings with competing, complex demands for attention). Thus, the ecological validity problem limits the validity of standardized assessments as predictors of real world outcomes.

Assessing Employability and 21st Century Skills

Gauging employability and 21st century skills that combine cognitive and non-cognitive domains, particularly aspects of collaboration, communication, or problem solving is vital. Technology can help in the construction of assessments that better resemble the contexts in which skills are required. This may include leveraging simulations, games, and/ or agent-based modalities to better approximate real-world demands. For example, to measure collaboration, traditionally highly qualified assessors would observe students in a simulation designed to evoke teamwork. While this ideal is achievable, it is also prohibitively expensive to conduct at scale. Costly human endeavors happen to be where technological solutions can assist in providing substantial efficiencies; we recognize that a technology-based replacement of a human-operated assessment environment may not be a fully equivalent replacement, but what we give up in genuine representation we may recover in cost efficiencies and in the richness of other peripheral and physiological data.

Illustrative of this kind of technology for assessment use is the work on SimCityEDU by GlassLab, Pearson and Educational Testing Service (ETS), which deploys an environmental simulation game where students find solutions to increasingly complex environmental challenges while being assessed on their ability to problem solve. Another example is the work on stealth assessment by researchers at Florida State University. Stealth assessment, in the context of game-based assessment, means not just tracking a player's score on the game, but also a player's progress on unobtrusively embedded assessment items on specific competencies, such as the ability to resist distractions, or to quickly identify patterns.

For low-income countries, RTI International has been researching the use of short, tablet-based games to assess employability skills such as task completion, time management, and problem identification. Most non-cognitive skills, unlike academic competencies in reading or mathematics, are inherently difficult to quantify, and thus their measurement has often relied on self-reported or trainer-reported questionnaires which are limited in the degree to which they provide evidence of the skills in question. While assessing skills in a simulated or mixed-reality environment is still only an approximation of the real world, these technology-mediated assessments make it possible to measure a student's application of these skills in a range of scenarios. This approach, in theory, should present a more nuanced understanding not only of the degree to which students exhibit these skills, but also under what conditions.

Formative Assessment

In the above examples, we have predominantly described the use of technology for summative or one-time student assessments. Technology can also facilitate continuous assessment, personalized learning, and the strategic selection of assessment items. This makes it possible to track individual student performance over time, and provide data-informed instructional guidance. In the United States, there are numerous such technologies already available, especially for assessments in reading, writing, and mathematics. In this U.S. case, the technology offers potential for instruction that bridges the gap between what individual students currently know and what they are expected to know. While it may be ambitious to expect teachers, especially where capacity is low and classes are large, to implement personalized instruction, such technology does support instructional decisionmaking in pacing, grouping, and material use, thus responding to students' needs in consideration of the curriculum. To date, however, there are few published examples of the use of technology for formative assessment in low-income contexts.



Identifying Disabilities

Furthermore, recent advances in electronic disability screening platforms for mobile devices are opening opportunities for more comprehensive assessments and understanding of learner capacities and needs. In Ethiopia, RTI has leveraged mobile phone-based technology for vision and hearing screening of over 3,700 students in combination with an electronic assessment of their reading skills. In another example, researchers at the University of Jyväskylä in Finland are examining the feasibility of a phonological awareness game to predict student dyslexia. In the United States, a team of researchers at the University of North Texas are investigating the use of Virtual Reality to assess neurocognitive skills and deficits, including supervisory attentional processing and Attention-Deficit/Hyperactivity Disorder. Technology thus facilities the integration of a range of assessment methodologies, providing a more differentiated and holistic understanding of a child's capabilities and needs.

POSSIBILITIES AND LIMITATIONS

Not every assessment situation lends itself to the use of technology, however. A recent reading and mathematics assessment done under the U.K. Department for International Development-funded Girls Education Challenge in Afghanistan reverted to the use of paper and pencil since electricity supply for technology-assisted data collection was impossible to guarantee.

The deployment of technology-mediated, multi-modal assessment approaches combining cognitive assessments with physiological data, may also not be feasible at scale in many settings. Instead, for the next few years, using such methodologies could inform the design and improvement of the assessment (e.g. its length), as well as related intervention programs, through relatively small sample-based applications. Eventually, such technologies could be used for universal measurement.

Similarly, the use of technology to facilitate sustained personalized assessment and content provision at the child level, is likely to remain unaffordable in low-income contexts at least for the next five to 10 years. Low-cost tablets cost around US\$40, an insignificant amount for high-income countries that spend upward of \$15,000 per student per year. But this is likely too high in countries like Malawi, which spend only \$30 per primary student per year.

The key questions to answer in using technology for assessment, include: What is the added value of the technology? What kind of interference will potentially come from use of the technology for assessment? How does introducing technology change what we are measuring? What may be possible ethical or privacy concerns in relation to the use of technology for assessment?

Furthermore, with the powerful presence of social media, we leave a trace not only of our "likes" and (implicit) dislikes, and therefore of our psychological makeup and its correlates, but, importantly from an educational perspective, we leave a trace of our cognitive skills and preferences. Facebook postings could be analyzed for our writing skills, what we search for in Google could be analyzed for the sophistication, so to speak, of our concerns. Short cognitive tests, common now in Facebook as a fun activity, could morph into the opportunity to literally test users for hours and hours and over time. This is likely to happen without the permission of governments, think tanks, academics, or NGOs. In fact, it is already happening. The issue, then, is that, since education remains a public good, how should the public view the use of social media technology in operating educational assessment platforms? While all this may seem futuristic to the poorest 40 percent of the world, current efforts by large tech companies to expand the internet to such segments of the population is profit-driven, and, given the technological might and deep pockets of these companies, is likely to happen before too long.

CONCLUSION

The use of technology for assessment purposes has great potential. Some of it, such as improving data collection and reporting, is easy to realize and is a worthwhile investment even in low-income countries, as the per capita cost of the technology is minimal if millions of children can be reached. Other uses, such as giving children individual tablets for personalized assessments and learning, may be out of reach for some time, especially in countries where the total yearly expenditure per child is lower than the cost of a tablet and capacity for sustained technology support is minimal. There are various "sweet spots" in the middle of these opposites. Ultimately, we believe that technology that improves the authenticity and validity of assessments of students' complex skills, or that provides a more holistic understanding of students' capabilities and needs, is worth pursuing.



ENDNOTES

- 1 Care, Esther, and Alvin Vista. "Education is Changing—It's Time Assessment Caught Up." SS/R. March 27, 2017. https://ssir.org/articles/entry/education_is_changingits_time_assessment_caught_up.
- 2 Care, Esther, Helyn Kim, Kate Anderson, and Emily Gustafsson-Wright. "Skills for a Changing World: National Perspectives and the Global Movement." *Brookings* Institution. March 28, 2017. https://www.brookings.edu/ research/skills-for-a-changing-world-2/.



Luis Crouch

Chief Technical Officer, International Development Group, RTI International

Luis Crouch, PhD, is a recognized international leader in providing high-level advice to governments involved in complex educational systems change. From 2011 to 2013, Dr. Crouch served with the World Bank's Global Partnership for Education Secretariat as head of the Global Good Practices Team.

He currently leads work addressing important challenges in education, workforce and youth, and "Data Revolution for Development." He provides input and oversight to key areas of work in all of the International Development Group's themes.

Dr. Crouch is also researching fundamental issues at the leading edge of applied scientific work on education while continuing to pursue his policy advisory work with specific countries in areas such as school funding and educational decentralization.





Carmen Strigel

Director, ICT for Education and Training, RTI International

Carmen Strigel's work in RTI's International Development Group strengthens efforts in areas such as education, workforce development, governance, policy making, stakeholder collaboration, and capacity building by leveraging the potential of Information and Communicity Technology (ICT). Ms. Strigel has extensive country experience in Africa, Asia, and the Pacific.

In the field of ICT for Education and Training, Ms. Strigel works on teacher training and professional development, e-content and instructional design, methodological/pedagogic integration of ICT into teaching and learning, organizational development and change management, and ICT policies and strategies, with a focus on enhancing instruction and student learning opportunities.

Ms. Strigel designed Tangerine®, acutting-edge, flexible, open-source electronic data collection software geared specifically for Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) and optimized for mobile devices. Ms. Strigel has researched, designed, implemented and monitored educational technology initiatives for more than 15 years for the European Union, United Nations Development Program, Asian Development Bank, and the United States Agency for International Development, among others.