

Assessment: Foundations and the Future

by Barbara Rowan, Ph.D. and Kate Parker

Assessment has long been a cornerstone of higher education. With the rise of online learning, the role of assessment is increasingly central. Advances in technology, including possibilities emerging from generative AI, have changed how assessments are created, delivered, and evaluated. New research has also changed the way we think about assessment, including a greater focus on how learners learn and how we can best measure learning accurately and equitably.

Here we provide a brief overview of the assessment field to ground us in the basics of assessment and share some key pointers to keep in mind as you determine how best to design and deploy assessment to support your learners' success.

Moments of Assessment

The assessment strategy for a course should be determined in the early stages of program and course design. You might begin by asking yourself these questions:

- Who are our learners, and how might we best support their persistence and engagement?
- How might our assessment strategy align with our goal of supporting a diverse, equitable, and inclusive learning experience?
- How will the course be delivered? Online or in person? Synchronously or asynchronously?
- What is the level of the course? Is this an introductory course or higher level?
- What do we want learners to know and be able to do at the completion of this course? Are there any industry or credential-specific requirements we need to design toward?

The answers to these questions will not only drive instruction but will also inform the assessment strategy.

When designing assessment, it is critical to understand not only *whom* you are assessing, *what* you are assessing, and *why* you are assessing it, but also *when* and *how* you are assessing learning. Assessment can play three fundamental roles in learning:

1. Assessment *for* Learning – Assessment *for* learning is what we often refer to as Formative Assessment. Here, assessment happens during the learning process, enabling learners to understand what they are to learn and what the expectations are. Learners are given feedback to improve their work and enable them to learn. Teachers use formative assessment to align instruction and identify specific learning needs. These are assessments such as knowledge checks embedded within an online course, daily exit tickets, or classroom quizzes. Formative assessments

can be auto-graded or manually graded and are meant to be less time-consuming than summative assessments – it should only take a few minutes for a learner to complete them (Bennett, 2017; Hall, 2020; Manitoba, n.d.; Sabtiawan et al., 2019; Swaffield, 2011).

2. Assessment of Learning – Assessment of learning (or of competence) is often referred to as Summative Assessment. Most often, these are given at the end of a lesson or unit and are used to provide evidence of student learning or competence. Summative assessment is usually higher stakes; that is, important decisions are often made based on summative assessment results (pass to the next level; pass to the next grade; be admitted to a special program; etc.). These assessments include unit tests, midterm exams, final exams, performance-based assessments, and many standardized tests (Bennett, 2017; Hall, 2020; Manitoba, n.d.).

3. Assessment as Learning – Assessment as learning takes the form of self-assessment or self-reflection. Learners need the opportunity to become metacognitively aware of how they learn. In assessment as learning, learners reflect on their own learning and make adjustments so that they can achieve deeper understanding. Reflection should be done regularly. Teachers should model and teach the skills for self-assessment. Teachers should also assist learners in setting their own learning goals and monitoring their progress toward them. Effective feedback is required in order to challenge ideas and introduce additional information. The learner focuses on the task at hand rather than getting a correct answer (Hall, 2020; Manitoba, n.d.; Sabtiawan et al., 2019).

In the sections that follow, we will focus on assessment of learning and assessment as learning. What does the research tell us about what works and why? How might we approach the design and development of these assessments more holistically?

Assessment of Learning: A New Approach to Linear Testing

In educational assessment, linear testing, also called fixed form or static assessment, is the most commonly used model. This model of assessment refers to the process of giving each student an identical test that contains exactly the same questions. Whether given on the computer or via paper and pencil, there are often only one or two forms of the test. If there is more than one form, there are a few questions in common (i.e., linking questions), so we can ensure the two forms are equivalent (Becker & Bergstrom, 2013; Frohn, 2021; Morton, 2018; Thompson, 2018).

There are issues associated with the linear assessment model. First, students repeating the test will likely receive exactly the same test they took the first time. It is likely that scores will be higher on the second attempt due to practice effects. The student has already seen all the questions (Kingsbury et al., 2009).

Another issue is that the test is more easily compromised in the linear assessment model. Because the same questions are used over and over, there is an increased opportunity that the test questions will be released on the internet. Students can share them with other students or with people outside of the university. And since there are a finite number of questions, the compromised test is no longer a good measure of learning (Becker & Bergstrom, 2013; Frohn, 2021; Kingsbury et al., 2009; Morton, 2018).

A different approach to linear testing is **linear-on-the-fly testing (LOFT)**. With LOFT, students take completely different versions of a test. The purpose of LOFT is to deliver a unique test to a student, but one that is psychometrically equivalent to every other form of the test. LOFT is typically used for summative and higher stakes assessment, so it falls under the assessment of learning role (Becker & Bergstrom, 2013; Frohn, 2021; Kingsbury et al., 2009; Morton, 2018; Thompson, 2018).

Within the LOFT framework, questions are created to build a sufficiently large item pool. Each question is tagged with metadata that could include the topic, learning outcome, competency, or whatever identifiers you want to use. Questions can be marked as *friends* or *enemies*, as well (Frohn, 2021; Morton, 2018).

Then, the test questions need to be given to at least 100 test takers in order to gather the psychometric data. The item data becomes part of the question metadata. The success of the LOFT model of assessment relies heavily on the quality of the question metadata, so always triple check that before deploying the test (Morton, 2018).

Once your questions are calibrated, you are ready to deploy LOFT. Whether students are testing at the same time or on demand, a unique test is created for each student. Questions are pulled based on your blueprint (e.g., 10 questions from each of the 5 domains) and any other metadata that you want considered, but also by the psychometric data. And you can be sure that friendly items are shown on the same form, but enemy items are not. An algorithm ensures that the overall test for each student is equivalent in terms of content and difficulty (Becker & Bergstrom, 2013; Frohn, 2021; Morton, 2018; Thompson, 2018).

Because each student receives a unique exam, it is unlikely that the test would be compromised. In the event that a form is compromised, you could potentially identify the individual who released the test questions, since no other student should have had this exact form. Further, you can remove compromised questions from the pool without affecting future forms of the test (Frohn, 2021; Morton, 2018).

The LOFT assessment model offers so much more over the traditional linear model of assessment. With increased security and unique but equivalent forms for each student, you can have faith in your measurement of student learning.

Assessment of Learning: Artificial Intelligence and Assessment

With so many rapid developments in generative AI emerging in recent months, it's hard not to get excited about its potential uses in assessment. It promises to save us time and money, but will it? What is AI in this context? And should we all sign up? Let's take a deeper look at the history here and the current state of the field.

The term *artificial intelligence* was first used by John McCarthy in 1955. Artificial intelligence refers to things that can simulate the human mind, including software or machines. AI can perceive patterns and problem-solve. It can communicate and learn. It's also pretty good at abstract reasoning (Caspari-Sadeghi, 2002; Cope et al., 2021; Shute & Zapata-Rivera, 2008). We've seen AI in self-driving cars, factory robots, smart technologies, and aircraft without human pilots. It takes a variety of technology types to deliver AI. And machine learning is what is used to learn and adapt the models we deliver (Caspari-Sadeghi, 2022; Cope et al., 2021).

Machine learning takes in information and uses statistical methods to create models for predictions that are based on observed patterns. Assessment is an area of opportunity for machine learning. Sure, AI and machine learning have been around for a while. And we have definitely used it in assessment. But there are now so many new opportunities for using AI and machine learning within educational assessment (Caspari-Sadeghi, 2022; Cope et al., 2021).

The Current Landscape

The use of AI in assessment has been around for a while. It's been used for scoring writing assignments and even for tutoring learners. But the call for AI assessment development has gone out, and it is being explored in many different ways.

Machine learning AI is fairly prevalent in education today. Its use in assessment typically fills the assessment *of* learning role, since it is primarily developed for summative and higher stakes assessment. Standardized testing companies have long since been using natural language processing to score essays. Educational technology companies followed suit so they could allow unlimited participation in their massive online open courses (MOOCs). Many states are also using natural language processing to evaluate the essay section in the end-of-year assessments (Gardner et al., 2020; Jiminez & Boser, 2021).

But now, machine learning promises us the answers to faster, better assessment development. And some are using it successfully.

Machine learning is promising for delivering situational judgment assessments. Chatbot conversations lead learners from one step to the next, and the system gives the next situation based on the learner's previous decision. The conversations and decisions are algorithm-based and determined by the learner's selections (Hosking, 2022).

AI/machine learning is used in patient simulations, fire-fighter training, and emergency medical technician programs. It's also used to introduce gaming into educational assessments (CareerCert, n.d.; Winkler-Schwartz et al., 2019). Further, it has been found to be effective in training engineering students using multidimensional quantitative and qualitative factors in the assessments, which provides individualized, personal feedback to the learners (Samarakou et al., 2014).

The Model Is Only as Good as Its Input

The thing about machine learning is that it is always learning. And it is this learning process that gives the machine the data it needs to build algorithms. These algorithms are used for making predictions and for tightening the model. The more good data you put in, the stronger the model (Caspari-Sadeghi, 2023).

Educators would love to have a system that automatically generates quality assessment questions. This is what AI can deliver to us – but only after the algorithm has been strengthened and the model is strong. To that end, it could take some time (and a whole lot of data) to get to the point where the model is creating assessment questions that follow best practices (Jiminez & Boser, 2021).

So, When Will We Get There?

AI and machine learning offer greater measurement precision and efficiency. It also can decrease bias (Hosking, 2022). But when will it be all we need it to be?

There is still a separation between education and technology. Educators and assessment specialists know where they want to go, but they lack the machine learning knowledge to get there. In addition, many schools and colleges lack the technological resources needed to develop the assessments and create the models. Further, exam developers need to have a strong data management plan, considering the risks associated with privacy, data security, fairness, and biased algorithms (Caspari-Sadeghi, 2023).

Finding a way to use AI to integrate instruction and assessment results could improve performance on educational outcomes. The potential is there. We just need the expertise and resources to get us over the finish line (Caspari-Sadeghi, 2023; Shute & Zapata-Rivera, 2008)

Assessment as Learning: Authentic Assessment and Metacognition

Instructors and instructional designers often face the challenge of how to make assessment more meaningful. How do we get beyond just grading an assessment to something more engaging with richer feedback than just a score? This is a strong case for introducing authentic assessment.

Odds are, you've heard the phrase "Authentic Assessment" a lot lately. It's considered best practice for authentically measuring learning. But what does that really mean? Let's take a closer look.

The term *authentic assessment* has been around since the early 1990s. At the time, it referred to direct assessment of knowledge – did the student learn what she was meant to learn? This direct assessment of knowledge was demonstrated through a project or an essay and was initially meant to be an alternative to the many standardized assessments that were being used in schools to measure learning (such as the Iowa Test of Basic Skills or the California Achievement Test). This direct assessment focused on what the student could **do** with their knowledge, a difference from answering multiple-choice questions to show what they can recall. And although more time consuming for the teacher to develop and score, the results would be more rich and meaningful (Fulton et al., 2021; McArthur, 2021; Raynault et al., 2022; Thomas, 2021).

If you talk to people who work in education today, it is very likely that they would say that authentic assessment refers to assignments that model **real-world experiences**. These could include assignments like writing a business plan, analyzing an accounting spreadsheet, writing computer code, or developing a website. These project-based assessments measure authentic knowledge, as opposed to the information recall often measured by multiple-choice tests (Abrams & Gerber, 2013; Fulton et al., 2021; Huang & Shideler, 2021; McArthur, 2022; Thomas, 2021).

Even in the literature, you will see reference to more **hands-on** types of authentic assessments used in education today. Some mention being more creative in assessment development, such as giving speeches, conducting research, or writing (Thomas, 2021). Many others say that authentic assessment allows students to simulate tasks they will likely encounter outside of the university classroom (Fulton et al., 2021).

But used to its fullest, authentic assessment extends beyond just project-based work. It's a holistic approach where teaching, learning, and assessment are woven together and occurring at the same time (Azim & Khan, 2012). And while hands-on, real-world tasks are so important for student learning, we also need to allow the student to focus on *how* she learns. Authentic assessment includes activities like **journaling** – ask the students to reflect on how motivated they are to start the next project or have them self-reflect on how they feel they learn best. Or, the journal prompt could ask the student to describe each step of a process used to complete an assignment. **Portfolios** are also an important authentic assessment tool. Not only can portfolios contain artifacts that may help the student get a job after college, but it can also show the learner how much he has learned over time (Ozan, 2019; Raynault et al., 2022). Other forms of authentic

assessment include **collaboration** among students, **self-evaluation** of projects, and **peer evaluations** of assessments (Raynault et al., 2022; Ozan, 2019).

In authentic assessment, **feedback** is crucial – feedback from instructors *and* peers. By feedback, we don't mean praise. Instead, students need very specific feedback to point out how the student has done thus far, and, most importantly, where they should go next. Instructors should be intentional about detailing why a learner earned the score they did. What could the learner have done differently to improve their outcome? What can the learner do in the future to improve their performance? Instructors must also be certain the student can read the feedback and understand what they mean.

Self-evaluation and **peer feedback** should be utilized, as well. As mentioned earlier, self-evaluation can come in the form of journaling or completing a rubric to evaluate one's own project. Peer feedback can also come from rubrics used to evaluate other students' projects (Dawson et al., 2021; Mandouit, 2016). There are many established benefits of self-evaluation and peer feedback. Self-evaluation provides skills for evaluating one's work with an eye toward how to improve it. Self-evaluation can also be used to make learners more aware of considerations related to diversity, equity, and inclusion. (Tessera, n.d.; Harvard University, n.d.; University of Delaware Center for Teaching & Assessment of Learning, n.d.).

With peer feedback, learners develop skills in providing feedback to others. Through the process of peer feedback, the peer reviewer not only evaluates another learner's project but also sees how others approached the assignment. The act of reviewing the work of others can result in greater learning for the reviewer. The recipient of the peer feedback can gain valuable formative information before the instructor evaluates the project. This feedback goes a long way toward bolstering the learner's confidence in the end product's quality (Larkin, 2014; Wood & Kurzel, 2008). Peer feedback can also help sensitize learners to bias and stereotypes. Peer reviewers learn to provide respectful feedback and become more aware of other cultures, ethnicities, and gender identities. When instructors normalize peer reviews within a DEI framework, learners see the world in a different way (Tessera, n.d.; Yarmey, n.d.). Providing peer feedback through a DEI lens can help learners think critically and creatively across all learners' attributes and abilities (University of Delaware Center for Teaching & Assessment of Learning, n.d.).

Fortunately, authentic assessment is here to stay. Research consistently demonstrates that authentic assessments provide strong measurement of real-world tasks. They are also a more meaningful measure of what students can do. At a time when encouraging lifelong learning is emerging as increasingly critical, authentic assessment enables students to learn more about themselves and understand how they learn (Ozan, 2019; Sabtiawan et al., 2019; Thomas, 2021).

About the Authors

Dr. Barbara Rowan is the owner of All Things Research, LLC, a company that provides assessment, psychometric, data analysis, and research consultation and practice to schools, colleges, and corporations. Currently, she is serving as an assessment consultant with LearningMate.

Before starting her own business, Barbara held various assessment and research positions at publishing and educational companies. She teaches graduate-level courses in assessment, research methods, organizational behavior, and statistics, both on-ground and online.

Barbara holds a B.A. in mathematics and English from Malone University, an M.S. in psychology from Georgia College & State University, and a Ph.D. in assessment and measurement from James Madison University in Harrisonburg, VA.

Kate Parker is Vice President of Content Services for Higher Education at LearningMate. In this role, she collaborates with publishers and institutional clients to design and develop engaging learning experiences that align to outcomes, competencies, and workforce-relevant skills. Kate has led this business for LearningMate for over six years, in which time her team has designed and developed thousands of courses and assessments, serving hundreds of thousands of learners worldwide. Prior to joining LearningMate, Kate spent over 15 years leading the sciences publishing program at Macmillan Learning.

Kate holds a B.A. in English Literature from Reed College and certificates in human-centered design from IDEO U.

Interested in learning more?

At LearningMate, our experts in assessment and instructional design collaborate with publishers and institutions to design and develop engaging learning experiences at scale. Please contact us: communications@learningmate.com to find out how we may be of service to your team.

References

- Abrams, S. S., & Gerber, H. R. (2013). Achieving through the feedback loop: Videogames, authentic assessment, and meaningful learning. *English Journal*, 103(1), 95-103.
- Becker, K. A., & Bergstrom, B. A. (2013). Test administration models. *Practical Assessment, Research & Evaluation*, 18(14). <http://pareonline.net/getvn.asp?v=18&n=14>
- Bennett, J. (2017). Assessment for learning vs. assessment of learning. *Pearson*.
<https://www.pearsonassessments.com/professional-assessments/blog-webinars/blog/2017/12/assessment-for-learning-vs--assessment-of-learning.html>
- Caspari-Sadeghi, S. (2023). Artificial intelligence in technology-enhanced assessment: A survey of machine learning. *Journal of Educational Technology Systems*, 51(3), 372-386.
- CareerCert. (n.d.). *How AI and technology are revolutionizing firefighting*. Retrieved April 14, 2023, from
<https://www.careercert.com/blog/fire/how-ai-and-advanced-technologies-are-revolutionizing-firefighting/>
- Cope, B., Kalantzis, M., & Searsmith, D. (2021). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*, 53(12), 1229-1245.
- Dawson, P., Carless, D., & Lee, P. P. W. (2021). Authentic feedback: Supporting learners to engage in disciplinary feedback practices. *Assessment & Evaluation in Higher Education*, 46(2), 286-296.
- Frohn, S. (2021, May 25). *Increase security and efficiency with linear on the fly testing (LOFT)*. PSI.
<https://www.psiexams.com/knowledge-hub/increase-security-efficiency-linear-on-the-fly-testing-loft/>
- Fulton, J., Scott, P., Biggins, F., & Koutsoukos, C. A. (2021). Fear or favor: Student views on embedding authentic assessments in journalism education. *International Journal of Work-Integrated Learning*, 22(1), 58-71.
- Gardner, J., O'Leary, M., & Yuan, L. (2020). Artificial intelligence in educational assessment: "Breakthrough? Or buncombe and ballyhoo?" *Journal of Computer Assisted Learning*, 37, 1207-1216.

- Hall, W. (2020, December 10). *What is assessment as learning? Enhancing teaching with data*. Century. [https://www.century.tech/news/what-is-assessment-as-learning-enhancing-teaching-with-data/#:~:text=Assessment%20as%20Learning%20\(AaL\)&text=%E2%80%9CIt%20considers%20how%20pupils%20self,learn%20about%20themselves%20as%20learners.](https://www.century.tech/news/what-is-assessment-as-learning-enhancing-teaching-with-data/#:~:text=Assessment%20as%20Learning%20(AaL)&text=%E2%80%9CIt%20considers%20how%20pupils%20self,learn%20about%20themselves%20as%20learners.)
- Harvard University (n.d.). *Self-assessment: The DIEB Learning Tool Project*. Retrieved April 24, 2023, from <https://projects.iq.harvard.edu/deib-explorer/activities>
- Hosking, T. (2022, November 24). *5 ways talent acquisition technology is revolutionising the search for talent*. Talent.ai. <https://talent.ai/5-ways-talent-acquisition-technology-is-revolutionising/>
- Huang, S., & Shideler, A. (2021). Leveraging student strengths through project-based learning and authentic assessment in an integrated ENL classroom. *Journal for Leadership and Instruction*, xx, 42-46.
- Jimenez, L., & Boser, U. (2021). *Future of testing in education: Artificial intelligence*. The Center for American Progress.
- Kingsbury, G. G., Bontempo, B., & Zara, A. R. (2009). *A comparison of CAT with LOFT methods for certification examinations*. Presentation at the 2009 NOCA Annual Educational Conference.
- Larkin, T. L. (2014). The student conference: A model of authentic assessment. *International Journal of Engineering Pedagogy*, 4(2), 36-46.
- Mandouit, L. (2016, August 4). Authentic feedback: What it is and isn't. *Education Weekly*. <https://www.edweek.org/education/opinion-authentic-feedback-what-it-is-and-isnt/2016/08>.
- Manitoba. (n.d.). *Assessment and evaluation: The role of assessment in learning*. Retrieved on April 14, 2023, from <https://www.edu.gov.mb.ca/k12/assess/role.html#:~:text=Assessment%20for%2C%20as%20and%20of.embedded%20in%20the%20learning%20process.>
- McArthur, J. (2022). Rethinking authentic assessment: Work, well-being, and society. *Higher Education*, 85, 85-101.
- Morton, L. (2018, May 1). To LOFT or not to LOFT, that is the question. *TestReach*. <https://www.testreach.com/blog-post/To-LOFT-or-not-to-LOFT-Linear-on-the-Fly-Testing.html>

- Ozan, C. (2019). Authentic assessment increased academic achievement and attitude towards the educational measurement of prospective teachers. *International Journal of Evaluation and Research in Education*, 8(2), 299-312.
- Raynault, A., Heilporn, G., Mascarenhas, A., & Denis, C. (2022). Teaching experiences of e-authentic assessment: Lessons learned in higher education. *Journal of Teaching and Learning with Technology*, 11, 3-17.
- Sabtiawan, W. B., Yuanita, L., & Rahayu, Y. S. (2019). Effectiveness of authentic assessment: Performances, attitudes, and prohibitive factors. *Journal of Turkish Science Education*, 16(2), 156-175.
- Samarakou, M., Fylladitakis, E. D., Prentakis, P., & Athineos, S. (2014). Implementation of artificial intelligence assessment in engineering laboratory education. *International Conference e-Learning*, xx, 299-303.
- Shute, V. J., & Zapata-Rivera, D. (2008). Educational assessment using intelligent systems. ETS.
- Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433-449.
- Tessera, E. (n.d.). *How can we incorporate diversity, equity and inclusion in evaluation*. Eval Academy. Retrieved April 24, 2023, from <https://www.evalacademy.com/articles/how-can-we-incorporate-diversity-equity-and-inclusion-in-evaluation>
- Thomas, M. (2021). The everyday creativity of authentic classroom assessments. *LEARNing Landscapes*, 14(1), 393-407.
- Thompson, N. (2018, August 2). *Linear on the fly testing*. Assessment Systems Corporation (ASC). <https://assess.com/linear-on-the-fly-testing/>
- University of Delaware Center for Teaching & Assessment of Learning (n.d.). *Diversity and inclusive teaching*. Retrieved April 24, 2023, from <https://ctal.udel.edu/resources-2/inclusive-teaching/#:~:text=Diversity%20and%20inclusion%20improve%20teaching,abilities%20and%20attributes%20and%20embraced.>
- Winkler, Schwartz, A., Yilmaz, R., Bissonnette, V., Ledwos, N., Siyar, S., Azarnoush, H., Karlik, B., & Del Maestro, R. (2019, August 2). Machine learning identification of surgical and operative factors associated with surgical expertise in virtual reality simulation. *JAMA Network Open*, 2(8), e198363. <https://doi.org/10.1001/jamanetworkopen.2019.8363>

Wood, D., & Kurzel, F. (2008, January). Engaging students in reflective practice through a process of formative peer review and peer assessment. *ATN Assessment 08: Engaging Students with Assessment*.

Yarmey, N. (n.d.). *Evaluation has a racism problem - What can we do about it?* Eval Academy.
<https://www.evalacademy.com/articles/evaluation-has-a-racism-problem-what-can-we-do-about-it?rq=diversity>