The Common Core State Standards, the Next Generation Science Standards, and the Potential of Digital Game-Based Learning and Assessment

Eric Tucker, Design Innovation Factory
Erin Mote, Design Innovation Factory

Introduction

Over the past decade, game-based learning and assessment have emerged as promising areas of innovation that could inform and enhance the provision of personalized learning at scale. Perhaps the greatest influence of game-based experiences on the next generation of learning will be not as instructional courseware per se, but as assessments *for* learning—that is, assessments that not only measure learning, but also generate actionable evidence to help personalize and improve teaching and learning.

This paper presents promising practices in the emerging field of technology-enhanced, game-based educational assessment. We examine recent developments in the field and present interviews and case studies to illuminate potential paths along which game-based assessment might evolve. We aim to increase understanding of how game-based assessment could become an integral component of assessment practices and systems, and we encourage the game industry, the assessment industry, and "games for learning" advocates to develop game-based assessment products with the potential to enhance meaningful accountability and to inform teaching and learning processes. We seek to advance games, learning, and assessment as an emerging field and to suggest that game-based assessment may hold relevance for broader conversations about next generation assessments, particularly those aligned with the Common Core State Standards (CCSS) and the Next Generation Science Standards (NGSS). Finally, we generate recommendations for designing and applying digital games to address the demands of assessment.

The recommendations are based on applied research conducted by the authors. Over nine months, beginning in early 2013, we conducted either in-depth interviews or surveys with over 125 professionals with relevant expertise, including leaders from game companies, education technology firms, publishers, research universities, school systems, non-profits, entrepreneurs, advocacy organizations, and policy researchers. We also reviewed pertinent academic and policy research and industry reports on the CCSS' relevance for game-based courseware efficacy, design, and dissemination.

Assessment in a K-12 Context

The U.S. education system is no stranger to assessment. Increasingly, schools and teachers have been evaluated on the basis of their effectiveness at advancing student learning. No Child Left Behind (NCLB) ramped up attention to standardized testing by mandating assessment and accountability for all states. Now, the adoption of the CCSS for Mathematics and English Language Arts (ELA)/Literacy and of the NGSS has created a need for corresponding next generation assessments. By reconciling and aligning existing state standards, CCSS adoption offers an opportunity for systematic, strategic thinking about how we might redesign the assessment system in a way that is more conducive to teaching and learning. Coupled with new game-based learning technologies, CCSS adoption offers two major opportunities for assessment design:

- *Prospect 1:* Assessments that serve, inform, and enhance teaching and learning. Formative assessments have the potential to be effective instructional tools, but they have not been widely used because they are burdensome for teachers to create and implement during limited class time. However, CCSS expectations and new game-based learning technologies make formative assessments both desirable and feasible.

- *Prospect 2:* Assessments that leverage technological advances to more efficiently assess a wider range of competencies. The transition to digital assessment is underway, necessitated in part by the transition to the CCSS and NGSS. New assessments developed to align with the CCSS will be computer-based and leverage technology in a range of ways. These technology-enhanced assessments promise to allow assessment of a wider array of cognitive competencies and to increase the efficiency of assessment.

The current assessment system has been subject to frequently raised concerns, which may be relevant to designing more promising teaching- and learning-centered assessment:

- *Lack of precise, actionable information.* Traditional assessment does not provide the sort of precise, actionable information necessary to improve instruction and learning. This limited information makes current assessments inadequate as instructional tools.

- *Static snapshots.* Summative assessments were designed for comparability across schools and systems accountability purposes; they provide static snapshots of achievement at particular moments but fail to capture real-time progressions of student mastery that would allow teachers to adjust and improve instruction during the course of the year.

- *The design of reliable and valid measures.* One challenge in providing more precise, actionable, and comprehensive information is ensuring that measures of learning processes and outcomes are both reliable and valid.

- *Fairness and equity.* Beyond the issue of test bias, the use of assessments for high-stakes decisions raises broader issues of fairness: when high-stakes assessments are poorly aligned with the actual content taught to students, they are unfair proxies for student achievement. Unfair or inaccessible assessments cannot be used effectively to improve instruction because they are poor measures of actual student learning.

- *Test administration and interpretation.* Quality assessment depends on design and how tests are used and administered. Ensuring that schools and teachers administer and interpret tests properly entails careful consideration of how data should be interpreted, whether interpretations are meaningful, and what the stakes are.

The Next Generation of Assessments

The next generation of assessment should seek to move beyond the limitations of status quo assessment toward a new assessment system that exists in service of teaching and learning. The next generation of assessment needs to take seriously the separation of assessment and instruction—and the possibility of overcoming it—and look towards standards for quality, including comprehensiveness, coherence, continuity, and dynamism.

Comprehensiveness

Assessment systems should be comprehensive for educators and policymakers interpreting results, meaning that "a range of measurement approaches should be used to provide a variety of evidence to support educational decision-making" (Pellegrino, 2012, pp. 10-11). The range of measurement approaches that schools and teachers use should generate a variety of evidence and provide multiple pathways for students to demonstrate competence and mastery. A critical component of a comprehensive assessment system is formative assessment, which provides real-time feedback to inform teaching and learning.

Coherence

Assessment systems should be vertically coherent—from large-scale summative assessments to classroom-based formative assessments—and horizontally coherent with curriculum and instruction. If the underlying models of learning are consistent, assessments have the potential to complement each other rather than conflict as a student moves from the classroom through the school, district, and state. The CCSS provide a model of student learning that can create consistency across summative and formative assessments.

Continuity

An assessment system ideally provides continuous records of progress so that change over time can be observed and interpreted. Pellegrino (2012) compares continuous assessment to "a videotape record rather than to the snapshots provided by most current tests" (pp. 10-11). Continuous assessment requires "multiple sets of observations over time," which "must be linked conceptually so that change can be observed and interpreted. Models of student progress in learning should underlie the assessment system, and tests should be designed to provide information that maps back to the progress. Thus, continuity calls for alignment along the third dimension of time" (Pellegrino, 2012, pp. 10-11). Compared to snapshot summative assessments, continuous assessment systems are better able to assess "the processes of learning and an individual's progress through that process" (Pellegrino, Chudowsky, & Glaser, 2001, pp. 256-257).

Dynamism

Next generation assessments should be dynamic, meaning that they are able to diagnose and guide personalized learning and support interaction and complex learning tasks. Formative assessment and adaptive instruction could address "the unique and situation-specific needs of learners by concurrently providing clear information, opportunities for thoughtful practice, informative feedback, and a favorable combination of intrinsic and extrinsic motivators tailored to the individual learner" (Dieterle & Murray, 2009, p. 601). By leveraging technology, next generation assessments could enable self-assessment and engage students in learning while they complete assessment tasks.

Potential Added Value of Game-Based Assessment

Game-based assessments could offer additional sources of evidence to support educational practice and decision-making. In a coherent assessment system, game-based assessments would be grounded in the same underlying model of student learning as other assessments. Game-based assessment, like most technology-enhanced assessment, has several potential advantages. Games can accommodate complex approaches to questions and responses; embed formative assessments and offer timely feedback; provide rich insight into learning progressions; generate data to explain and improve learning outcomes; and give users control, agency, context, and motivation.

Complex approaches to questions and responses

Games can be an important part of next generation assessment systems that ask new, more complex questions with more flexible response options. Next generation assessments will "go well beyond traditional item formats" to ask students to "connect knowledge, processes, and strategies to conditions of use"; that is, next generation assessments must "make mandatory the use of more complex tasks, including simulations and other extended constructed-response formats" (Bennett, 2013, p. 127). For example, an assessment might ask students to engage in adaptive problem solving, in which they are given a situation and must "identify the problem and its constraints, represent the problem so that it can be solved, figure out alternative solution options, implement one strategy, and evaluate the adequacy of the solution or solutions" (Baker, 2012, pp. 10-11). In such a format, they cannot simply regurgitate recently learned procedures, but must "draw from previously learned patterns or schema, adapt them to the problem at hand, and perhaps invent something new in order to reach a solution" (Baker, 2012, pp. 10-11). The transition to next generation formats will be driven by "the need to measure competencies that cannot be assessed through less labor-intensive means" (Bennett, 2013, p. 127).

Embedded formative assessments and timely feedback

At present, it is impractical for many classroom teachers to administer robust daily or weekly formative assessments. Doing so requires significant time and attention, increasing demands for managing data collection, flow, analysis, and application. But game technology has the potential to ease the logistical demands and complexity of providing formative assessment feedback by making collection, interpretation, and reporting simpler and less time consuming. By embedding assessments, games could become "potent formative assessments tools" (Baker & Delacruz, 2012, p. 3). Data could be captured "through evaluation of students' online clickstream behavior to support inferences about students' ongoing understanding" (Baker & Delacruz, 2012, p. 3).

Rich insight into learning progressions

Recent technological advances have spawned what is now known as "big data." Our activity on the Internet, for example, can be analyzed to make predictions about our future behavior. Next generation game-based assessments can harness big data and provide us the means for gathering on-the-fly data about student progression, mastery, and learning, which can then be used to make instructional decisions tailored to individual needs. Stealth assessments, embedded in immersive game environments, could be used to continuously monitor performance and give feedback.

Data to examine learning progressions and mastery and improve teaching

By monitoring students as they engage in the process of learning, game-based assessments can help refine learning progressions to more accurately guide instruction (Pellegrino, 2012, p. 5). Games can enrich the problems posed by assessments; they can present complex, realistic, open-ended, interactive challenges; and generate evidence about the approaches students choose. Baker (2012) calls this "one of the biggest potentials for game design—the ability of process data to help explain learning outcomes (e.g., use of productive or unproductive strategies), sense and adapt to students' evolving understanding of the domain, misconceptions, or gaps in domain

knowledge" (p. 3). Behrens and DiCerbo (2012) embrace this capability and suggest "a reframing of assessment practices from identifying correctness of test questions to capturing a constellation of learning transactions using digital technologies to make inferences about student cognition and learning" (as cited in Gordon Commission, 2013, p. 20).

Self-assessment

Digital games can facilitate the shift toward self-assessment by making assessment personal, accessible, and immediate. Self-assessment is a significant dimension of next generation learning. It is a crucial skill for developing motivated learners who are able to take stock of where they are and what they need to improve. In other words, self-assessment can "encourage students to reflect about what they've learned and how well they've internalized and understand it; how it all comes together; how it has or can change their behavior; and what else is needed to continuously achieve their own as well as society's goals" (Torre & Sampson, 2012, p. 6). It helps to fill the gaps between school- or teacher-led assessments and ensure that learning happens continuously.

Validity through engagement and motivation

Incorporating game-based elements into assessment may bring the added advantage of engaging students. For instance, adaptive instructional and assessment tools are already transforming basic math instruction, through tools such as DreamBox learning and Dragon Box. Engagement is not only about student experience; it also helps ensure assessment validity. Research suggests that less engaging assessment tasks tend to underestimate student abilities (Bauer et al., 2012, p. 24). Especially in literacy assessments, research finds, achievement scores may depend on "providing interesting text, providing clear conceptual goals, providing choice, and collaboration"—properties that are "uniquely suited to, or can be more readily effectively incorporated within, technology enhanced assessment formats" (Bauer et al., 2012, p. 24). By increasing assessments' novel interactivity, and by providing immediate, task-relevant feedback, technology can increase assessment validity (Bauer et al., 2012, pp. 24-25). Because games are so engaging, they can also extend students' time-on-task by migrating learning opportunities beyond formal educational settings. This advantage may be especially relevant for students at risk of academic failure (Baker & Delacruz, 2012, p. 1).

Opportunities for Games, Learning, and Assessment

Develop game-based interim and formative assessments

Game-based assessment should include interim (periodically administered) and formative assessments that will work together with large-scale, summative assessments in a system that is comprehensive, coherent, continuous, and dynamic. Game-based assessment would thus be one type of evidence to support educational practice and decision-making and would be grounded in the same underlying model of student learning outlined by the CCSS and NGSS. Research and development of new digital courseware and assessments are required to understand how, if, and to what extent game-based assessment systems can come together with assessment systems currently in place.

Align digital games with the CCSS and NGSS

To achieve their full potential, learning games should be designed from the ground up to be integrated with instruction and assessment. Game-based assessment should not be designed to exist in isolation, but should be aligned with the CCSS, the NGSS, and the assessment consortia. The CCSS and NGSS signal worthwhile goals for schools, educators, and students. The aspects of learning that game-based systems assess and emphasize should aim to be consistent with the evolution of classroom practice. The CCSS and NGSS call for meaningful, worthwhile work—meaning that "teaching and learning to the test" has the potential to become a good thing. From the outset, designers should "design the assessment architecture, in other words, embed the assessment in the transactions of the game and build it into a game's underlying engine. Maximizing the potential of process data requires a tight (conceptual) coupling among the set of goals of learning, purpose of assessment, student behaviors, student responses, task design, and assessment design" (Baker & Delacruz, 2012, p. 3).

Common standards also help preserve equity: "common understandings and common standards for performance for both accountability and instructional purposes are required if equity is to be served and performance disparities reduced. If students in different schools are held to vastly different types of performance, equity issues will exponentially increase with performance assessment" (Baker, 2010, p. 8). There is a need for increased and sustained collaboration focused on both the theoretical and practical matters of game-based assessment and its relevance to the assessment consortia. This applies not only to collaboration between researchers in game design and the

measurement sciences, but also to the collaboration of these groups with game and user interface designers, learning scientists, usability researchers, teachers, curriculum specialists, and assessment developers.

Produce game-based measurement tools that are fair, equitable, valid, and reliable

Game-based assessments have the potential to enhance fairness and equity and to improve validity. In order to evaluate validity, developers must first define the desired intellectual processes or academic outcomes to be measured as well as how the assessment will be used. Well-defined, common standards help clarify assessment tasks and help teachers understand the instructional tasks that will prepare students for assessments, further strengthening validity (Baker, 2010, pp. 6-7). Game-based assessment should value the need for high-quality evidence of CCSS skills mastery, including the need to meet our nation's measurement standards, including validity, reliability, fairness, and accessibility.

Conclusion

Together, technological advances and the adoption of shared next generation standards offer an opportunity to revisit and improve the existing assessment system. An improved assessment system would serve, inform, and enhance teaching and learning and leverage technological advances to more efficiently assess a wider range of competencies. Next generation assessment can move beyond the current system of snapshot assessments by developing a comprehensive, coherent, continuous, and dynamic assessment system with a robust formative assessment component. Game-based assessments are well suited to meet many of the needs of a next generation assessment system: they can accommodate complex approaches to questions and responses, embed formative assessments with timely feedback, generate data for rich insight, and encourage self-assessment and student engagement.

Realizing these opportunities will require the collaboration of policymakers, game designers, educational practitioners, investors, and many others. Our recommendations suggest ways in which each of these groups can support the development of game-based interim and formative assessments that are aligned with the CCSS, relevant to the broader assessment context, and based on Evidence Centered Design.

References

Baker, E. (2012). The future of learning and assessment in a changing world. Proceedings from National Center for Research on Evaluation, Standards, and Student Testing 2012: *First World Future Education Symposium*. Retrieved from http://www.cse.ucla.edu/downloads/files/Baker.Beijing.China%281%29.pdf

Baker, E. (2010). *What probably works in alternative assessment* (CRESST Report 772). National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from http://www.cse.ucla.edu/products/reports/R772.pdf.

Baker, E., & Delacruz, G. (2012). *What do we know about assessment in games?* National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from

http://www.cse.ucla.edu/products/overheads/AERA2008/baker_assessment.pdf.

Bauer, M., Marquez, E., Keehner, M., Laitusis, C., van Rijn, P., Zapata-Rivera, D., & Hakkinen, M. T. (2012). *Technology enhanced assessments in mathematics and beyond: Strengths, challenges, and future directions*. Proceedings from K-12 Center at ETS, 2012: *Invitational Research Symposium on Technology Enhanced Assessments*. Retrieved from http://www.k12center.org/rsc/pdf/session1-cayton-hodges-keehner-laitusis-marquez-paper-tea2012.pdf

Behrens, J. T. & DiCerbo, K. E. (2013). *Technological implications for assessment ecosystems: Opportunities for digital technology to advance assessment*. The Gordon Commission on the Future of Assessment in Education. Retrieved from http://www.gordoncommission.org/rsc/pdf/behrens_dicerbo_technological_implications_assessment.pdf.

Bennett, R. E. (2013). Preparing for the future: What educational assessment must do. In The Gordon Commission on the Future of Assessment in Education, *To assess, to teach, to learn: A vision for the future of education*. 123-141).

Dieterle, E., & Murray, J. (2009). Realizing adaptive instruction: The convergence of learning, instruction, and assessment. In Schmorrow, D. D., Estabrooke, I. V., & Grootjen, M. (Eds.) *Foundations of augmented*

cognition. Neuroergonomics and operational neuroscience, San Diego, CA: Springer. 601-610.

Gordon Commission on the Future of Assessment in Education. (2013). *To assess, to teach, to learn: A vision for the future of assessment*: *Technical report.* Retrieved from http://www.gordoncommission.org/rsc/pdfs/gordon_commission_technical_report.pdf

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C.: National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=10019.

Pellegrino, J. (2012). Assessment of science learning: Living in interesting times. *Journal of research in Science Testing*, 49(6).

Torre, C. A., & Sampson, M. R. (2012). *Toward a culture of educational assessment in daily life*. The Gordon Commission on the Future of Assessment in Education. Retrieved from http://www.gordoncommission.org/rsc/pdf/torre_sampson_toward_culture_educational_assessment.pdf.