The Promise of Scenario-Based Assessments for College Instruction

Jonathan D. Cohen, Omer Ari, Emily Bryans Dobar, Gertrude Tinker Sachs, John Sabatini, Tenaha O'Reilly, & Joseph P. Magliano

Assessment-based Learning | Higher Education | scenario-based assessment

The lecture-high stakes test structure of many college courses does not promote deep learning that can be transferred to real world contexts. Scenario-based assessments (SBAs) provide an approach to assessment that grounds authentic problem solving and scaffolds students to a solution for those problems. The paper reports the first iteration of a design-based research (DBR) design cycle to develop an SBA authoring system, which will enable college instructors to use SBAs in their courses. An SBA was administered in an introduction to interdisciplinarity course taught at a large, public university. Data from students and faculty suggest that it was usable, and they perceived it to support learning beyond traditional course assessment formats. We reflect upon the affordances of SBAs that led to these positive perceptions.

Introduction

Many college courses follow a structure that is arguably not optimal with respect to promoting deep learning that can transfer beyond the classroom. Instructors in many disciplinary-specific courses (e.g., psychology, biology, history, etc.) structure their courses around a series of lectures interspersed with high-stakes assessments. Moreover, many of these tests adopt formats (e.g., multiple choice tests), that are problematic because they (a) are not grounded in theories of learning (Mislevy & Yan, 2017), (b) are decontextualized (O'Reilly et al., 2018), and (c) discourage transfer (Foster & Piacentini, 2023).

Students deserve better. Scenario-based assessments (SBAs) provide a potential solution to this systemic dilemma (Sabatini et al., 2020). SBAs are designed to afford students opportunities to apply the knowledge and skills they acquire in their courses. They ask students to address or solve broad, purpose-driven problems in ecologically valid settings. A well-designed SBA walks students through a task sequence that models strategies used by experts, while providing valid and reliable measurement of critical skills. SBAs typically have the following characteristics (e.g., Sabatini et al., 2020): (a) they explicitly model the critical steps in a complex performance task; (b) they ground that task in the context of solving a real-world problem; (c) the task sequence simulates task-appropriate strategies, improving student engagement and providing opportunities to develop critical metacognition and self-regulation skills; and (d) individual tasks (or items) measure critical supporting skills (or key supporting knowledge) needed to carry out specific steps in the task sequence.

The current study is the first iteration of a design-based research (DBR) design cycle (Reeves, 2006) intended to lead to an SBA authoring system for college instructors. DBR parallels learning engineering (Baker et al., 2022) in that each involves iterative, rigorous, data-driven analysis of theoretically informed interventions to support the development of learning environments. We designed and implemented an exemplar SBA in order to better understand both the process of designing SBAs and the role the SBAs played in the students' learning. It was guided by the following research questions: (a) What was the perceived usability of the SBA? (b) What are the student perceptions of the SBA? and (c) What are the instructor's perceptions of the SBA? To answer these research questions, we employed a concurrent mixed methods design, leveraging data collected from students and the instructor of an introduction to interdisciplinary thinking course in Spring 2025.

Methods

Participants, Course Context, and the Scenario-Based Assessment

Nine students enrolled in an introduction to interdisciplinarity course at a large, public university participated in the study. The university is an urban, primarily minority serving institution. The participants resembled the university's students in general, with most students (66%) between 18 and 24 years old. Forty-eight percent of university students identify as Black or African American, 22% as White, 21% as Hispanic. Nearly half (46%) of students are first-generation college students.

Interdisciplinarity in Human Learning and Development is a course in which students learn about the concept of interdisciplinarity and interdisciplinary thinking. Repko et al.'s (2019) textbook, Introduction to Interdisciplinary Studies, is used as the core reading. It introduces students to the foundational knowledge of interdisciplinarity, then it introduces the Broad Model of the Interdisciplinary Research Process (IRP). The Broad Model involves six steps, from defining the problem to a literature search, and reflection on the problem. The course culminates with a paper in which students apply the model to a societal problem of significant personal interest, following the criteria specified for each of the steps.

The SBA situates students in a scenario in which they are college-aged interns at Comm360, a fictional community-based nonprofit. They are part of a team of interns, tasked with developing a structure for a report to the CEO of Comm360 describing how the nonprofit might address the issue of high school dropouts. The digitally delivered SBA is broken into screens that parallel the Broad Model. As students progress through the screens, they participate in activities such as

highlighting tasks, article summarization, and other constructed responses. The students' work follows them through the SBA, so their responses on activities form the foundation for subsequent activities.

Procedure

We collected data from the following sources: (a) a usability questionnaire delivered following the SBA implementation; (b) student artifacts, including their interactions with and responses to the SBA and reflections on the SBA process; and (c) instructor interviews.

We employed reflexive thematic analysis to analyze the qualitative data (Braun & Clarke, 2006), in which we collaborated through the process, developing a shared interpretation of the data. For this study, a group of five researchers, all of whom are grant personnel and including the course's instructor, participated in the analysis. The first step was familiarization with the data, followed by the first phase of inductive coding. From these codes, we generated a list of themes, which we subsequently discussed, refined, and named. These themes focused our continued exploration of the data. We used basic descriptive statistics to analyze the usability questionnaire data. The questionnaire is a modified version of the System Usability Scale (Brooke, 1996), which is a 10-item questionnaire, widely used for this type of evaluation. Our modified version adapted 9 of the items for our study. It has the advantages of being brief and can be used with small sample sizes, but lacks in fine-grained, formative diagnosis of specific design issues.

Results

What Is the Perceived Usability of the SBA?

Of the consenting students, 11 completed the usability questionnaire. Participants rated 8 of the 9 survey items at or above a 4.00 on a 5-point Likert scale (1, Strongly Disagree to 5, Strongly Agree), suggesting that, overall, participants felt positively about the usability of the SBA. Most of the items explored students' perceived usability of and confidence with the SBA interface, but the first item was I felt confident in my understanding of the content to gauge students' comfort with the content of the SBA. This item received an average rating of 4.18, suggesting 10 of the 11 students felt confident with the content. With an average rating of 4.27, 10 of the 11 participants also indicated agreement with I found the content to be consistent with what we were learning in the course.

The highest-rated item was I think that I would need the support of a technical person to be able to use the SBA interface. After reverse-coding this item since it was negatively worded, the average rating was 4.64 out of 5 (1.91 before reverse-coding), suggesting 9 of the 11 students indicated they could easily use the interface without external support. The lowest-rated item received an average rating of 3.73, I think that I would like more SBAs to be used in other courses, a result which will be explored in future iterations of the SBA.

What Are the Student Perceptions of the SBA?

An overarching theme derived from the students' reflections is that their experience applying the Broad Model was much messier than they anticipated. The weeks leading up to their participation in the SBA were spent learning the model as linear, sequential steps in a process, as is introduced in the textbook. The SBA, though, simulated the application of the Broad Model in an authentic setting, and the nature of that simulation required a more recursive application of the Broad Model. This shift in understanding was a revelatory moment for some of the students. Some correctly perceived that their previous conceptions of the model as linear were more a function of its presentation in the textbook, and their reflections demonstrated an understanding of the need to modify the process to react to circumstance. For example, one step of the model involves explicitly justifying an interdisciplinary approach. The scenario they were presented with elided this step, as one student

described: “In the SBA we never really had to do that, our ‘supervisor’ was immediately on board.” Other students experienced the recursive application of the model similarly, noting that this was a change in their previous understanding of the model.

Another experience noted by the students was their interactions with their simulated colleagues (i.e., dialogic agents) within the SBA. The simulation was minimal; the “colleagues” were simply images and text, and there was no real interactivity between them and the students. However, some students reflected that even that minimal level of simulation of collaboration supported their work. This was evident in the use of the first-person plural in many of the students’ reflections when referencing their action steps during the SBA (e.g., “our discussion”, “our team at the non-profit”, etc.). Some students explicitly referenced the agents as important parts of the learning experience. One noted that the introductions to their “colleagues” at the beginning of the SBA “made me think it was real classmates”. This connection appeared to be a motivating factor for some of the students, who referenced the “realness” of the teamwork when describing completing the SBA.

In addition to the motivational aspect of the agents, some students reflected on how the scaffolding from these digital teammates were important parts of their SBA experience. For instance, one remarked that “having the other interns’ answers available allowed me to check and compare my answers.” Another discussed how the agents’ perspectives “broaden[ed] my horizons past where I would’ve originally looked at.” Clearly the students did not believe their SBA colleagues were actual people, but the students seemed to suspend disbelief while working through the scenario, allowing for a more “real” seeming environment. This realness was a constant theme in the students’ reflections.

What Are the Instructor Perceptions of the SBA?

The instructor for this course has seven years of experience teaching the Broad Model. Watching his students use the model over the years, the instructor noted that the step-by-step application of the model often fell short of the potential for promoting interdisciplinary thinking. He observed that students tended to check off steps performatively. Some of this is a consequence of the prescriptive nature of some steps of the model. For example, one step asks students to declare that the problem is complex, resulting in students simply writing “The problem is complex.” The order of some steps also did not make sense. In one of the earliest steps, students were asked to state that there was more than one discipline offering insights into the problem before they had actually reviewed the literature and confirmed that was the case. Implementing the IRP steps in this format felt like busywork, leaving students without the rich insights they were supposed to gain from following the model. He characterized the work as often not “cognitively engaged” and that it felt “uninspired.”

This contrasted with students’ performance on the SBA, which the instructor described as richer and deeper. As he put it, “the SBA brought the Broad Model to life,” as the steps were applied authentically within the simulated context of a nonprofit committee tasked with developing a more comprehensive understanding of the causes of the high school dropout problem in the community they serve. Embedded in the natural-sounding dialogic interactions, the steps were applied purposefully and authentically to the work of the committee. Students were more engaged in this more fluid interdisciplinary thinking process, which incorporated the same six steps of the Broad Model of IRP. He indicated that students were more invested in the topic and wanted to continue working towards a solution, with one student commenting, “We evaluated different perspectives but never got to the problem-solving portion.”

Discussion

This study demonstrated that students found the SBA usable, and in general, the students’ reflections and the instructor’s perceptions were consistent in reporting that students developed a more nuanced and realistic understanding of how the Broad Model can be used to solve problems that are interdisciplinary in nature. The SBA format afforded the designers opportunities to create a structure which helped students achieve the learning objectives. In particular, a primary affordance of the SBA is its ability to approximate an authentic setting within which to situate students’ work. Certainly, part of that authenticity stems from having images to represent the agents and including language that sounds truer to the types of

individuals students could reasonably expect to encounter in the scenario. Beyond these more superficial elements, the authenticity stems from situating the SBA in a real-world environment that demands fluidity and complexity. The way models are presented in textbooks are flat representations, which push students toward a flatter understanding of processes. SBAs afford situated application of knowledge and skills rather than simple recitation and therefore help students to develop a more nuanced understanding of those knowledge and skills (e.g., Bransford, 1997).

The SBA format also affords designers the ability to model behaviors and skills related to the learning objectives. In this SBA, the interactions between the students and the dialogic agents provided opportunities to model the application of the steps of the Broad Model. Research demonstrates that dialogic agents support learning, even in a context when they are not truly interactive with students (Discoll et al., 2003). It was through the modeling and scaffolding provided through the dialogic agents that students were led to understand the steps of the Broad Model are less linear and more recursive than they perceived them to be after reading the textbook.

It is also notable that this SBA was implemented in an asynchronous online environment. Such environments can feel isolating for students (Croft et al., 2010), as student-student and student-instructor interactions are digitally mediated and temporally separated in asynchronous modalities. The presence of agents–even minimally interactive ones, as found in this SBA–was a motivating factor for the students.

Scenarios form the backbone of SBAs and afford designers the ability to integrate sequenced activities which can scaffold cognitive processes. At the beginning of this SBA, the scenario provided an authentic opportunity to activate students' prior knowledge. Later parts of the SBA attached targeted literacy activities (i.e., finding and highlighting relevant text) to the narrative. The SBA concluded with activities asking the students to synthesize ideas they generated throughout the SBA. In real-world situations where students actually work at a community non-profit, they could reasonably expect to do similar activities, adding to the perceived authenticity of the experience. Moreover, the structure of the scenario narrative forces designers to consider a range of pedagogical moves and scaffolding that fit into the narrative, as the scenario will feel flat and inauthentic if they do not.

In conclusion, this study suggests that SBAs can be applied to post-secondary instructional contexts. Their affordances may allow college instructors to break from the traditional lecture-high states exam structure of college courses and develop a context that can help students learn that they can transfer what they learn to real-world problems.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305T240021 to The University of Memphis. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Baker, R. S., Boser, U., & Snow, E. L. (2022). Learning engineering: A view on where the field is at, where it's going, and the research needed. Technology, Mind, and Behavior, 3(1), 1–23. https://doi.org/10.1037/tmb0000058

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. Measurement: Interdisciplinary Research and Perspectives, 8(2–3), 70–91. https://doi.org/10.1080/15366367.2010.508686

Bransford, J. D. (1997). The Jasper project: Lessons in curriculum, instruction, assessment, and professional development. Routledge.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

Brooke, J. (1996). SUS: a “quick and dirty” usability scale. In P. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), Usability Evaluation in Industry (pp. 189–195). CRC Press.

Croft, N. Dalton, A. & Grant, M. (2010) Overcoming isolation in distance learning: Building a learning community through time and space. Journal for Education in the Built Environment, 5:1, 27-64, DOI: 10.11120/jebe.2010.05010027

Driscoll, D. M., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. C. (2003-2004). Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. Journal of Educational Computing Research, 29(4), 431–450. https://doi.org/10.2190/Q8CM-FH7L-6HJU-DT9W

Foster, N., & Piacentini, M. (2023). Innovating Assessments to Measure and Support Complex Skills. Organisation for Economic Co-operation and Development.

Mislevy, R. J., & Yan, D. (2017). Evidence-centered assessment design and probability-based inference to support the generalized intelligent framework for tutoring (GIFT). In R. A. Sottilare, A. C. Graesser, X. Hu, & G. A. Goodwin (Eds.), Design Recommendations for Intelligent Tutoring Systems: Volume 5—Assessment Methods (Vol. 5, pp. 101–124). U.S. Army Research Laboratory.

O’Reilly, T., Sabatini, J., & Wang, Z. (2018). Using scenario-based assessments to measure deep learning. In K. K. Millis, D. Long, J. P. Magliano, & K. Wiemer (Eds.), Deep Comprehension: Multi-disciplinary Approaches to Understanding, Enhancing, and Measuring Comprehension. (1st ed., pp. 197–208). Routledge.

Reeves, T. (2006). Design research from a technology perspective. In J. Van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), Educational Design Research. Routledge.

Repko, A. F., Szostak, R., & Buchberger, M. P. (2019). Introduction to Interdisciplinary Studies. SAGE Publications.

Sabatini, J., O’Reilly, T., Weeks, J., & Wang, Z. (2020). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. International Journal of Testing, 20(1), 1–23. https://doi.org/10.1080/15305058.2018.1551224