

A Learning-oriented Assessment Perspective on Scenario-based Assessment

Brian A. Carroll

Teachers College, Columbia University

INTRODUCTION

When most people think of language assessments, they generally recall the standardized formats of high-stakes, large-scale examinations. These assessments are widely recognized for providing valid and reliable measures of test-taker knowledge. However, recently these standardized assessments have also been criticized for not providing accurate reflections of the knowledge, skills, and abilities essential for 21st century students to confront the era of digital communication. In order to meet these demands, assessment researchers and test designers have been re-conceptualizing test design and development procedures to provide more authentic assessments which more accurately reflect the complexity of tasks test-takers are likely to encounter in the 21st century (Bachman & Palmer, 1996; Chapelle & Douglas, 2006; Purpura, 2004).

Solutions: Innovative Frameworks and Guidelines

Learning-oriented assessment (LOA) is theorized as a development and validation framework for identifying the dynamic, interactive relationships between instruction, learning, and assessment in the classroom. Aside from examining classroom-based dynamics, this framework also recognizes the influence of educational standards and technology on the development of curriculum and materials used in both education and assessment (Purpura & Turner, forthcoming). The orientation to learning in LOA prioritizes the use of positive and negative feedback on performance to moderate the relationships between learners and learning processes and the indicators of learning demonstrated by performance. The LOA framework also highlights the inter-relationships between instruction, learning, and assessment to both foster successful learning and guide the development of learning-oriented assessments.

Educational Testing Service developed the Global Integrated Scenario-based Assessment (GISA) suite of assessments to reflect advancements in cognitive and learning sciences to meet contemporary expectations for 21st century students (Sabatini, Halderman, O'Reilly, & Weeks, 2016). The GISA development and design procedures and methodologies are aimed at capturing test-taker ability to employ higher-order thinking skills and engage in purposeful, 21st century assessment activities (Sabatini, O'Reilly, Halderman, & Bruce, 2014). Emerging from GISA, scenario-based assessment (SBA) has become a standardized set of guidelines used to develop assessments across several subject areas with as many as twenty-four different test-formats over K-12 grade levels (O'Reilly, Deane, & Sabatini, 2015; Sabatini, Petscher, O'Reilly, & Truckenmiller, 2015). SBA incorporates sequences of thematically related test-tasks promoting knowledge and skill development through the use of scaffolding and feedback. This sequence is situated within an interactive, web-based storyline presenting test-takers with

a series of simulated social interactions and test-tasks directed at accomplishing a single, overarching goal representative of 21st century activities (Sabatini et al., 2014).

Unfortunately, to date, no formal analysis has attempted to integrate SBA development and design guidelines into the LOA theoretical framework. An analysis of the SBA guidelines through the lens of the LOA framework will unify these two emerging perspectives and, in so doing, indicate the extent the theoretical rationales of the LOA framework can support the SBA guidelines. Such an examination can both specify standard principles and practices for developing and designing language assessments to meet 21st century needs and reveal the potential future direction of language assessment development and design.

EXAMINING SBA THROUGH LOA

LOA conceptualizes the dynamic interactions between instruction, learning, and assessment along seven dimensions classified under two main categories, referred to as performance moderators and performance indicators. While the five dimensions in the former category indicate the influences affecting learning development, the two dimensions in the latter category identify methods for eliciting, evaluating and interpreting the success of learning development. The following sections will indicate specific features of SBA guidelines according to the multidimensional framework of LOA.

Performance Moderators

The LOA framework identifies the extent five separate yet interacting performance moderators, referred to as context, instruction, interaction, affect, and socio-cognitive demands interact to mediate learning development. Similarly, SBA guidelines also conceptualize performance moderators; however, here they are considered assessable cognitive behaviors--rather than theoretical dimensions--which influence the development of knowledge and skills. SBA identifies these moderators as background knowledge, level of engagement, level of motivation, use of metacognition and self-regulation, and the use of cognitive strategies. Despite similar considerations, LOA and SBA share slightly different perspectives: The LOA framework theorizes these cognitive behaviors as characteristics of five separate, interacting dimensions, whereas SBA views performance moderators as assessable factors with the potential to indicate points of disruption in knowledge and skill development (Sabatini et al., 2015).

Contextual dimension

The contextual dimension of LOA identifies the effect of political and educational perspectives on the social, cultural, and technological factors related to learning development and performance. Related to this and fundamental to the motivation behind the development of SBA, Sabatini and O'Reilly (2016) address these contextual factors as the basis for expanding upon existing theoretical constructs in order to develop and

design assessments representing a full range of purposeful, 21st century activities. In their view, the rapid evolution of the role of technology in society--largely dictated by cultural views--influences the behaviors and performances of 21st century students and as such, should guide the development of contemporary theoretical constructs. While not specifically addressed in SBA research, the consequences of re-conceptualizing constructs underlying behavior are far-reaching; political and educational institutions utilize these construct definitions as standards to guide learning and learning outcomes for making decisions and allocating resources. Despite not specifically recognizing the potential consequences of re-defining construct definitions utilized in assessments, SBA research does acknowledge several principles of the contextual dimension guiding the development of scenario-based assessments.

Instructional dimension

In order to facilitate the management and organization of assessments, instructors must investigate the extent they effectively leverage technology, content knowledge, and pedagogical content knowledge to guide learning development and performance. These considerations under the instructional dimension are evident in the development of the *Cognitively Based Assessment of, for and as Learning* (CBAL), which provided much of the initial research for SBA. The development of CBAL was largely informed by researching specific instructional practices found in classrooms across subjects such as the English language arts, mathematics, and sciences. This research identified and generalized instructional practices into assessment procedures targeting test-takers' abilities to develop and integrate knowledge and skills (Bennett & Gitomer, 2009; O'Reilly, Deane, & Sabatini, 2015). This assessment approach enables the identification of partial knowledge or incomplete skill development as a means to support and guide further instruction. Additionally, within the assessments themselves, SBA examines the effect prior content knowledge, or background knowledge, has on moderating the ability to further develop knowledge and skills (Sabatini et al., 2014; Sabatini et al., 2016). Scenario-based assessments typically begin with probing this background knowledge, and, as the test progresses, participants review and integrate new information to further elaborate upon their existing knowledge. At the completion of the assessment, test-takers are evaluated for the extent the new information has been integrated into prior knowledge to form a more complex mental representation of that knowledge. At the developmental and within-test level, SBA works to reveal the relationships between content and pedagogical content and the development of knowledge and skills.

Socio-cognitive dimension

SBA guidelines address several features of the socio-cognitive dimension, which promotes the identification, comprehension, and integration of learning targets and examines the cognitive, socio-cognitive and strategic demands of assessing such learning targets. SBA assesses a test-taker's ability to use cognitive and metacognitive strategies--such as memory, transfer of knowledge, and self-regulation--to evaluate and synthesize thematically related sources of information to form mental representations of that information, which test-takers then use to generalize rules to solve novel problems

(Sabatini et al., 2014). One SBA test-format of GISA presents test-takers with several, web-based sources of information based on organic farming, and, as the test progresses, the test-takers are assessed on their cognitive and metacognitive ability to evaluate and synthesize sources to perform a culminating task, in which they design a website directed at teaching organic farming procedures and techniques. The outcome of assessing cognitive and metacognitive strategies and skills lends insight to potential disruptions in the development of knowledge and skills supporting knowledge development.

An additional feature of the socio-cognitive dimension found in SBA uses in-task feedback as a means to further knowledge and skill development. For example, as an innovation in assessment design, the organic-farming test-format utilizes virtual, simulated characters to give positive, motivational in-task feedback to promote learning-strategy management. In this test-format, when students answer questions they either receive praise from their virtual peers for correct answers, or, in the case of incorrect answers, a virtual teacher allows them to revise their knowledge and prove they have acquired the essential knowledge or skills. This structure and sequence of tasks and feedback support the identification and remediation of potential disruptions in knowledge or skill development. The socio-cognitive dimension highlights several features of SBA, which are considered essential aspects of SBA assessment development and design procedures.

Interactional dimension

The LOA framework uses conversation analytic (CA) procedures to investigate interactional structures embedded within classroom-based assessments. As mentioned above, SBA sequences virtual, simulated interactions to promote the development and integration of knowledge and skills, and to identify disruptions or gaps in knowledge and skill development by providing guidance and scaffolding to further test-takers' processing and accomplishment of task goals throughout the assessment. However, LOA conceptualizes several features of CA--such as turn taking, preference structure, and repair--not mentioned in the SBA literature. While SBA guidelines do not explicitly acknowledge such social and cultural interactional norms mediating virtual, simulated interactions, tacit knowledge of cultural norms enables the design of interactional sequences throughout the assessment. While these interactions may lack theoretical support, these sequences of virtual, simulated interaction are important to the development of SBA as they guide and scaffold knowledge and skill development throughout the assessment.

Affective dimension

This dimension of the LOA framework examines the socio-psychological impact of assessments on the attitudes and beliefs of test-takers and those who use assessment information to make decisions. SBA uses the simulated, virtual interactions to not only give feedback and guidance but also support an overall friendly, supportive assessment environment in order to reduce test anxiety and act as affect-moderating devices throughout the assessment. In researching the added value of the virtual, simulated peers, SBA researchers identified a positive relationship between the amount of time students

spend interacting with the virtual characters and their overall success on the assessments (Sabatini et al., 2014). Additionally, SBA guidelines also recommend incorporating fun and engaging themes and tasks, and using cartoons and humor to maintain student engagement and moderate affect throughout assessments (Sabatini et al., 2016).

In sum, the five dimensions of LOA performance moderators offer questions and theoretical support for better understanding the interrelationship between several SBA design principles moderating proficiency and performance. However, a complete analysis of SBA guidelines must also address several points beyond surface level design features and examine the underlying features of the assessments, or the indicators of performance—such as the development of the underlying theoretical constructs of proficiency, the characteristics of the elicitation of performances, and the evidence generated by those performances.

Performance Indicators

Referred to as performance indicators, LOA theorizes a series of multi-dimensional, dynamic interactions which guide and support decisions related to performance elicitation methods, the interpretations of evidence gathered from those methods, and the extent these decisions and interpretations are influenced by evolving notions of proficiency across levels and over time. The following section will discuss the methods SBA utilizes to elicit, evaluate, and interpret examinee performance, and examine the ways in which SBA test-tasks support assessment of knowledge and skill development.

Elicitation and evidence dimension

Considerations under the elicitation and evidence dimension are related to the design, organization and structure of test-tasks for eliciting specific types of evidence evaluated and used for both decision-making and promoting further knowledge and skill development. Associated with these considerations, SBA guidelines propose structuring and sequencing test-tasks in a manner which enables assessments to examine not only the result of knowledge development but also the acquisition of the skills supporting the development of knowledge. In addition to identifying potential disruptions in knowledge and skill development, the test structure and sequence also utilizes assistance and feedback as a means to overcome these disruptions. Different SBA test-formats, ranging across several topics and grade levels, have structured and sequenced test-tasks according to the developmental stages identified in empirical and educational research. For example, one SBA test-format, related to the topic of dolphin intelligence, designed for 6th grade English language arts students, presents test-takers with several different reading tasks (O'Reilly, Deane, & Sabatini, 2015). The tasks assessing reading comprehension sequence a series of sub-tasks which assess the stages of knowledge and skill development identified in cognitive sciences: identifying relevant information, extracting main ideas, defining vocabulary from context, and engaging in metacognitive behaviors to organize, produce, and evaluate summaries. Similarly, another SBA test-format, based on the topic of organic farming and designed for K-3rd and 6th graders, structures test-tasks to examine knowledge and skill development by assessing the ability

to organize content, comprehend details, evaluate web-based information, evaluate advantages and disadvantages, define vocabulary in context, evaluate the quality of differing opinions in online forums, and finally to summarize test content (Sabatini et al., 2014; Sabatini et al., 2016). These two examples showcase SBA considerations for the structure and sequence of test-tasks in relation to an understanding of the development of knowledge, and they investigate the manner in which these sequences can promote skill development.

Finally, while the previous examples relate mostly to evaluating student performance, this dimension is also concerned with the reliability and validity of assessments. One major challenge for SBA was to minimize test-fatigue and maintain test-taker engagement and focus while still providing a reliable measure of knowledge and skills within a single, 45-minute class period. A second challenge relates to SBA's 21st century construct definition, which requires test-takers to integrate multiple sources of information across different modes and within a single assessment. Traditionally, this raises issues related to item independence; however, the fact these tasks are sequenced and dependent upon one another actually enables the identification of points of disruption in learning processes. SBA research has published data confirming both the reliability and validity of SBA to measure knowledge and skill development within a single class period (Sabatini et al., 2014).

Proficiency dimension

The proficiency dimension of assessments examines the evolving conceptualizations of proficiency over time and considers the effect this evolution has on the display and evaluation of proficiency. In other words, a test should be constructed with: first, a consideration for the construct definition underlying proficiency; second, an examination of how this definition changes over time; and last, an understanding of the effect these changes have on test design. Throughout this analysis, SBA has been shown to assess both the result of knowledge development, as well as the development of the skills which underlie the further acquisition of knowledge. At another level, SBA guidelines have been used to develop at least two-dozen test-formats, ranging in complexity across proficiency levels in grades K-12, covering increasingly complex topics such as the life of a chicken, desertification, dolphin intelligence, space and satellites, and organic-farming. It is evident SBA considers both the extent test content is appropriate for assessing a wide-range of proficiencies across developmental stages, as well as whether test-task sequences can reveal the development of knowledge and skills over time.

CONCLUSION

A broad view of assessment development and design procedures portrays existing theories and methodologies as not fully representative of the knowledge, skills, and abilities students need to perform in the 21st century. Acknowledgement of these inadequacies has led to developments in assessment research attempting to meet contemporary expectations for student performance. LOA theorizes a framework of

dynamic, interacting relationships influencing instruction, learning, and assessment through which assessments can be developed, designed, and validated. As a complement, SBA, contains a suite of unique assessment development and design guidelines intended to reflect contemporary notions of education found in the cognitive and learning sciences. This analysis has indicated the extent to which the LOA framework can be a theoretical support underlying the SBA development and design guidelines conceived to confront 21st century expectations for assessments.

REFERENCES

- Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Bennett, R.E., & Gitomer, D.H. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt-Smith & J.J. Cumming (Eds.), *Educational assessment in the 21st century*. New York, NY: Springer.
- Chapelle, C.A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge, UK: Cambridge University Press.
- O'Reilly, T., Deane, P., & Sabatini, J. (2015). Building and sharing knowledge key practice: What do you know, what don't you know, what did you learn? *ETS Research Report Series, 15* (24), 1-53.
- Purpura, J.E. (2004). *Assessing grammar*. Cambridge, UK: Cambridge University Press.
- Purpura, J.E. (2014). *What is LOA?* Retrieved on 12/12/2017 from <http://www.tc.columbia.edu/tccrisls/what-is-loa/>
- Sabatini, J., & O'Reilly, T. (2013). Rationale for a new generation of reading comprehension assessments. In B. Miller, L.E. Cutting & P. McCardle (Eds.), *Unraveling reading comprehension: Behavioral, neurobiological and genetic components* (100-111). Baltimore: Paul H. Brooks Publishing Co., Inc.
- Sabatini, J.P., O'Reilly, T., Halderman, L., & Bruce, K. (2014). Broadening the scope of reading comprehension using scenario-based assessments: Preliminary findings and challenges. *L'Annee Psychologique, 114*(4), 693-723.
- Sabatini, J.P., O'Reilly, T., Halderman, L.K., & Bruce, K. (2014). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research & Practice, 29*(1), 36-43.
- Sabatini, J., Petscher, Y., O'Reilly, T., & Truckenmiller, A. (2015). Improving comprehension assessment for middle and high school students: Challenges and opportunities. In K. L. Santi and D.K. Reed (Eds.), *Improving reading comprehension of middle and high school students* (Vol. 10) (pp. 119-152). Springer International Publishing Switzerland.
- Sabatini, J.P., Halderman, L.K., O'Reilly, T., & Weeks, J.P. (2016). Assessing comprehension in kindergarten through third grade. *Topics in Language Disorders, 36*(4), 334-335.
- Turner, C.E., & Purpura, J.E. (forthcoming). Learning-oriented assessment in the classroom. In D. Tsagari & J. Banerjee (Eds.) *Handbook of second language assessment*. Boston, MA: DeGruyter Mouton.

Brian A. Carroll is currently a doctoral student in the Applied Linguistics Program at Teachers College, Columbia University, focusing on second language assessment. Additionally, Brian is a Community Language Program Teaching Fellow at Teachers College, Columbia University. Previously, he held the position Technology Fellow at Teachers College, Columbia University. He has taught EFL/ESL in St. Petersburg, Russia, New York, New York, and most recently Taipei, Taiwan. Brian also currently teaches ESL for the English Language Institute at City College of New York, CUNY. His main areas of interest include second language assessment, technology and neuro-cognition. Correspondence should be sent to Brian A. Carroll at bac2159@tc.columbia.edu.