Title of Project:
Investigating the Construct of Topical Knowledge in a Scenario-Based Assessment Designed to Simulate Real-Life Second Language Use

Researcher:
Heidi Liu Banerjee
Teachers College, Columbia University
heidi.liu@tc.columbia.edu

Research Supervisor:
Dr. James E. Purpura
Teachers College, Columbia University

Final Report

Motivation for the Research

The vast development of digital technology and the widespread use of social network platforms have reshaped how we live in the world. For second language (L2) learners to maximally utilize their language proficiency to function effectively as members of modern society, they need not only the necessary L2 knowledge, skills, and abilities (KSAs) but also essential topical knowledge. While many researchers believe that topical knowledge should be viewed as an integral component of L2 communicative competence, the role of topical knowledge has not always been accounted for in an assessment context due to the difficulty of operationalizing the construct.

Scenario-based assessment, an innovative, technology-based assessment approach, allows great affordances for expanding the measured constructs of an assessment. It is designed expressly for learners to demonstrate their KSAs in a context that simulates real-life language use. Through the utilization of a sequence of thematically-related tasks, along with simulated character interaction, scenario-based assessment offers opportunities to examine L2 learners’ communicative competence in a purposeful, interactive, and contextually meaningful manner.

Acknowledging the importance of including topical knowledge as part of the broadened construct of L2 proficiency, as well as the potential of utilizing scenario-based assessment to simulate the complexity of real-life language use, the primary purpose of this study was to investigate the relationships between topical knowledge and L2 KSAs in a scenario-based language assessment (SBLA). A set of language tasks surrounding a particular theme was designed to elicit L2 learners’ reading, listening, and writing abilities. Additionally, a topical knowledge task related to the same theme was designed as a performance moderator (O’Reilly & Sabatini, 2013) to activate prior knowledge, as well as to track topical learning. L2 learners’ various language and topical background characteristics were also explored to determine whether and to what extent they played a role in L2 learners’ demonstration of topical learning and overall L2 performance as measured by the SBLA.
Research Questions

The following research questions were addressed in the current study:

1. What is the relationship between L2 learners’ pre-scenario topical knowledge, post-scenario topical knowledge, and their L2 performance in an SBLA?
2. How well can the items in the topical knowledge task provide evidence for construct validity?
   a. To what extent do the topical knowledge items display adequate psychometric properties for their function as a measure of L2 learners’ topical knowledge?
   b. To what extent is there evidence in support of the use of the same set of topical knowledge items to track L2 learners’ topical learning?
3. To what extent do L2 learners with different language and topical background characteristics, specifically, L2 proficiency levels (intermediate, high-intermediate, advanced), fields of expertise (English, science, other), and degrees of prior personal experience with the topic (low, medium, high), vary in their topical learning and L2 performance?
4. In what ways are L2 learners’ abilities to achieve the scenario goal dependent on their topical knowledge and L2 KSAs? In other words, what inferences can be made about L2 learners who display certain levels of topical knowledge, listening and reading abilities, and their eventual success in achieving the scenario goal via summary writing?

Research Methodology

Context of the Study

The context of the current study, specifically for the assessment development, was set in the Community English Program (CEP) at a major research university in the United States. In order to place students entering the Program into appropriate course levels, a requisite placement exam is administered to all new CEP students at the beginning of every semester. Based on the results, incoming students are assigned to one of the 19 beginning, intermediate, or advanced levels the CEP offers. As an ongoing project, the CEP has been revamping its placement exam to better reflect the evolving construct of communicative language ability. The assessment instrument (i.e., the SBLA) developed in this study was designed for the high-intermediate level scenario component of the new CEP placement exam.

Participants

The participants in this study included 118 adult, Taiwanese English as a foreign language (EFL) learners, who served as test-takers, and two experienced English as a second language (ESL) teachers, who served as raters. In order to ensure that the characteristics of the participants were relevant to the research context, purposeful sampling (Wiersma & Jurs, 2009) was used. The EFL learners did not have direct connection to the CEP at the time of data collection; however, their proficiency levels were deemed appropriate to represent prospective test-takers who would be assigned to take the high-intermediate scenario module in the new
placement exam. Both of the raters have had extensive ESL teaching and testing experience in the United States.

**Instruments**

The instruments used in the current study included an SBLA titled “Nutrition Ambassador,” the SBLA experience survey, and the analytic scoring rubric developed to score test-takers’ responses to the writing task in the SBLA.

**Data Collection Procedures**

The test data were collected from EFL learners in Taiwan who voluntarily signed up to take the SBLA at a university computer lab. The SBLA was administered a total of six times over the span of three weekends. Two university students were hired as research assistants to proctor and monitor the test-taking process on-site, and the Researcher monitored the entire process remotely. The test-takers had a maximum of 110 minutes to finish the entire test. All of the test-takers finished the SBLA between 60 and 110 minutes. The assessment platform, Qualtrics, recorded and saved the test-takers’ responses automatically. All of the test-takers completed the test, and their responses were recorded successfully.

**Data Analysis Procedures**

The first research question, which investigated the relationships between topical knowledge and L2 performance within the SBLA, was examined by way of correlation and path analyses. Three hypothesized path models were analyzed to determine the effects between learners’ proficiency levels, topical knowledge, and L2 KSAs, as measured in this study.

The second research question, which investigated the extent to which the topical knowledge task in the SBLA could properly measure topical knowledge and track topical learning, was examined through a four-step Rasch analysis. The Rasch results were cross-compared to ensure that there was sufficient validity evidence in support of the intended use of the topical knowledge task.

The third research question, which investigated the roles L2 learners’ language and topical background characteristics (i.e., proficiency levels, fields of expertise, and degrees of prior personal experience with food additives) played in their topical learning and L2 performance, was examined through a series of one-way ANOVAs.

Finally, the fourth research question, which investigated the ways in which L2 learners’ ability to achieve the scenario goal (i.e., sharing knowledge of unsafe food additives) was dependent on their topical knowledge and L2 KSAs as measured in the preceding tasks within the SBLA, was examined using Bayesian networks.

**Summary of Findings**

Through correlation and path analyses, L2 learners’ pre-scenario topical knowledge and post-scenario topical knowledge were found to have different relationships with their L2 performance in the SBLA. The two aspects of topical knowledge as operationalized in this study, content knowledge (i.e., knowledge of factual information) and lexical knowledge (i.e., knowledge of lexical meanings), showed effects on L2 abilities differently; however, L2 learners’ proficiency level was found to account for their L2 performance the most.
Following that, validity evidence of the construct of topical knowledge as measured in the SBLA was provided using a four-step Rasch analysis. The results showed that the topical knowledge items were generally shown to fit the Rasch models well, demonstrating adequate psychometric properties for their functions as a measure of L2 learners’ topical knowledge. Because the same topical knowledge task was administered both in the beginning and at the end of the SBLA, the test-takers’ topical learning was tracked through the changes in item difficulty parameters. The estimated item difficulty change showed that, through contextualizing the knowledge building and sharing process, L2 learners were able to demonstrate substantial content learning. However, L2 learners did not seem to have learned the lexical items as much as they had the factual information of food additives. This is likely related to the ultimate goal of the SBLA: the test-takers were asked to share the information they had learned about food additives with their community. Therefore, the test-takers had to rely heavily on the content of the article in the reading task. It is important to note however, they were not asked to use the lexical items in any part of the assessment, nor was there explicit instruction of the lexical items during the knowledge building and sharing process. Such an assessment design may have contributed to the type of topical learning L2 learners demonstrated.

Then, taking a closer look at the roles L2 learners’ language and topical background characteristics played in their topical learning and L2 performance, a series of one-way ANOVAs was employed. The results revealed that L2 learners of different proficiency levels (intermediate, high-intermediate, and advanced) and fields of expertise (English, science, and other) varied in their content learning, lexical learning, and L2 performance. However, their prior personal experience with food additives did not appear to play a role here, suggesting that L2 learners’ self-identified life experience with a particular topic, which may be related to their episodic memory (Tulving, 1972), does not necessarily transform to readily accessible knowledge that can be utilized while the learners perform L2 tasks.

Finally, considering that the tasks in the SBLA were all thematically-related, a Bayesian network was constructed to examine the ways in which L2 learners’ ability to achieve the scenario goal depended on their topical knowledge and L2 KSAs. While the Bayesian network constructed in this study was exploratory in nature, and the sample size was not sufficient to yield robust generalizability of the results, it provided a holistic understanding of how L2 learners’ ability to achieve a communicative goal depended on their topical knowledge and L2 KSAs in the context of an SBLA. The results from Bayesian network also revealed that L2 learners’ ability to gain topical knowledge while completing a sequence of thematically-related L2 tasks appeared to be an essential part of their L2 communicative competence, and therefore, should be considered as a component of their “L2 proficiency score.” Lastly, with the increasing interest in adopting game-, scenario-, and simulation-based assessments to measure complex constructs of learners’ KSAs, this study demonstrated how Bayesian networks may be utilized to interpret the relationships between the measured constructs, so that results from these complex assessments can yield meaningful interpretations.

**Implications**

The current study carries a number of possible theoretical, methodological, and pedagogical implications for the field of applied linguistics, particularly in language assessment. Theoretically, this study contributes to the understanding of the role of L2 learners’ topical knowledge and its relation to L2 KSAs in a language assessment. In order to properly
capture the nature of topical knowledge, this study operationalized topical knowledge as both knowledge of topical content and knowledge of lexical meanings, and a set of items related to the theme of the scenario-based language assessment were designed to measure L2 learners’ topical knowledge. By administering the same set of items to the test-takers both before and after the scenario, it can be observed how much topical knowledge the test-takers already had (i.e., prior topical knowledge), and how much they learn about the topic in the process of achieving the scenario goal (i.e., topical learning). Through the test design, this study may provide insights into how L2 learners utilize their prior topical knowledge to complete the language tasks, and how a highly-contextualized scenario-based assessment may facilitate topical learning.

Methodologically, the current study informs the use of scenario-based assessment in L2 assessment contexts as well as the test design and the statistical procedure for an assessment with complex constructs. Because scenario-based assessment aims to simulate real-life language use, it is crucial for the storyline embedded in the scenarios to be coherent so that test-takers can perform in a way that is natural to their cognitive functioning (O’Reilly et al., 2015). With a coherent structure, the assessment results subsequently may allow test users to make meaningful interpretations of the evidence collected from test-takers’ performance. In order to do so, the scenario-based language assessment in this study adopts the key literacy practice of building and sharing knowledge as its theoretical framework, and an ECD framework as its design principle. The coherent test design also allows for the development of a Bayesian network to model the dependencies among L2 learners’ topical knowledge and their L2 KSAs, a measurement method rarely used in the context of L2 assessment. The scarcity of its use is primarily due to the fact that, until fairly recently, the technical constraints have made it difficult to measure complex constructs or simulate real-life language use coherently and systematically within an assessment.

Pedagogically, the current study attempts to address the facilitation of learning through meaningfully contextualizing an assessment, where test-takers can apply their L2 KSAs in an authentic manner (Hidalgo, Sata, & Suzuki, 2015). By examining the extent to which test-takers gained topical knowledge while completing the language tasks to fulfill the scenario goal, this study demonstrates how a purpose-driven, high-contextualized scenario-based language assessment could both be used to gauge L2 learners’ language proficiency and serve as a learning medium.
References


affordances of scenario-based assessment for broadening measurement opportunities in second or foreign language assessment”, Chicago, IL


