

Assessing Intercultural Understanding Through a Speaking Test

James Carpenter

Asia University

Sawako Matsugu

Rikkyo University

Reference Data:

Carpenter, J., & Matsugu, S. (2018). Assessing intercultural understanding through a speaking test. In P. Clements, A. Krause, & P. Bennett (Eds.), *Language teaching in a global age: Shaping the classroom, shaping the world*. Tokyo: JALT.

Study abroad programs represent an investment of time and money, for both students and universities. Therefore, assessing students' preparedness for studying abroad is important. Standardized English test scores are one way of identifying appropriate candidates. However, the relationship between scores and achievement from studying abroad is not well supported. There is a growing interest in assessing students' preparedness for studying abroad in a more holistic way. In this paper, a scoring rubric for an English language interview test-task is presented, which was designed to assess intercultural understanding. The theoretical justification for this test is that observable traits such as comfort level and engagement are as important as language ability for successful studying abroad. This paper first reviews previous efforts to assess intercultural understanding then presents the construct definition and an explanation of the scoring rubric. Finally, the administration of the test and future steps for validating similar tests are discussed.

大学における派遣型留学プログラムは大学と学生双方に時間と金銭的な投資を必要とする。それゆえに、いかに留学志望学生が留学に向けて準備できているか、ということの評価するのは重要である。外部英語試験のスコアの利用はそのような学生選定におけるひとつの手段である。しかし、それらの試験結果と留学における成功との関係ははっきりしていない。したがって、留学志望学生の選定において、より包括的な方法への関心が高まっている。本論文では留学志望学生を対象とする英語面接試験で使用した「国際理解」を評価するためのルーブリックを紹介する。これを開発した背景には、観察しうる快適度や従事度などの特性は、言語能力と同様に留学の成功のために重要だという認識がある。本論文はまず過去の関連文献を紹介し、次にルーブリックとその構成概念の説明をする。最後にテストの実施、そして今後の妥当性の検討方法を議論する。

Study abroad programs are a beneficial part of foreign language education. Such programs also represent a significant financial investment for universities. The return on this investment is connected, in part, to marketing successful study abroad programs to recruit new students. Long-term relationships with partner universities in foreign countries are also valuable from an administrative standpoint. Because of this combination of investment and the inherent responsibility universities have for educating students, it is in most universities' best interest not to encourage poorly prepared students to study abroad. Preparation for this naturally involves foreign language education. A minimum level of English ability is necessary for university students to study successfully in the United States. Many universities in Japan assess English skills by using commercially available standardized tests (e.g., Eiken and TOEFL) or a combination of test scores and more qualitative data such as statements of purpose, essays, or interviews. In many cases, the inclusion of such qualitative data when assessing a student's preparedness for studying abroad is preferable because English skills alone are not necessarily a good predictor of success overseas. Therefore, assessing students' suitability for studying abroad in a holistic way is important.

Intercultural Understanding

In this paper, it is proposed that assessing intercultural understanding is necessary to successfully evaluate students' preparedness for studying abroad. This is because the ability to function well in intercultural situations will necessarily enhance the study abroad experience. Celce-Murcia, Brinton, and Snow (2014) described the relationship between foreign language education and intercultural understanding in this way: "There is a renewed emphasis in education on teaching global citizenship, intercultural understanding, and lifelong learning and not only for the development of language proficiency across skill areas for immediate, local purposes" (p. 28). Intercultural understanding, in this context, may be understood as "the capability to be effective across different cultural contexts—including national, ethnic, generational, organizational, and

other contexts” (Livermore, 2013, p. 7). Successful intercultural experiences will, in turn, lead to future opportunities, and, potentially, to the above-mentioned increased interest in global citizenship and lifelong learning.

Intercultural Understanding in the Japanese University Context

The Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) emphasizes the importance of developing intercultural understanding; one current method for achieving this is to double the number of Japanese students studying overseas by the year 2020 (MEXT, 2017). Accomplishing this goal will depend in large part on factors outside of education and language policy. The Japanese population is shrinking. The Japanese Cabinet Office currently projects that the working population will decline by 30% over the next 40 years. This means that Japanese universities must compete for students as an increasingly scarce resource. In other words, there may be tension between Japanese universities’ political motivation to invest in study abroad programs (i.e., cooperating with the goals laid out by MEXT) and the financially driven motivation to manage study abroad programs in a conservative way.

Implicit in Celce-Murcia et al.’s (2014) definition is the idea that intercultural understanding is a democratic value that can, and should, be taught to everyone. Although Japan’s stated goals for English education are in line with the international teaching English to speakers of other languages (TESOL) movement, Japan’s social and economic reality may require a less ecumenical approach. In the face of a simultaneously decreasing student population on the one hand and a continued need for an internationally minded (and multilingual) workforce on the other, the question for many universities may be “which of our students has the greatest capacity for intercultural understanding *now*?” rather than “how can we teach intercultural understanding to as many students as possible?” Evaluating this requires well-designed modes of assessment.

Assessing Intercultural Understanding

Assessing intercultural understanding is not a simple task. There have been a variety of attempts to define the construct of intercultural understanding. Specific definitions have included (a) intercultural sensitivity, (b) mindfulness and attention to intercultural issues, (c) positive disposition towards other cultures, (d) willingness to engage, and (e) self-awareness (Sercu, 2010). Including such aspects in the construct presupposes test-tasks designed to elicit and evaluate attitudinal behaviors such as willingness to engage and interest in other cultures. In other words, intercultural

understanding is a separable language skill that can be evaluated using performance assessment (Liddicoat & Scarino, 2010).

As for any assessment, it is essential first to define the construct to be tested. A test construct is “the specific definition of an ability that provides the basis for a given assessment or assessment task and for interpreting scores derived from this task” (Bachman & Palmer, 2010, p. 42). The rubric for assessing intercultural understanding described in this paper is primarily task-based (see Appendix). This means that the criteria on the rubric are associated with real-world expectations and that a high score indicates a higher likelihood of real-world success (Brown, Hudson, Norris, & Bonk, 2002). In this case, a high score indicates a test-taker’s ability to function well in intercultural situations, and a low score indicates the opposite.

Test Background

The rubric described in this paper was developed for a newly created civil engineering department at a private university in Japan. This university has a large study abroad program and regularly sends students to countries throughout the world. The study abroad program with partner universities in the United States is the largest. Due to a variety of factors, the American study abroad program has faced financial and curricular difficulties. In addition, the civil engineering department places special emphasis on studying abroad in developing countries throughout Southeast Asia. As a result, the civil engineering faculty decided to only allow students with a score of at least 600 on the Test of English for International Communication (TOEIC), or a score of at least 400 on the TOEIC combined with a speaking test score designed on a 200-point scale, to study abroad in the United States. This requirement was imposed in part because the American study abroad program for the civil engineering department lasts approximately six months and includes an internship at an American-based Japanese company. The speaking test presented here was designed to assess students’ preparedness for this experience. The test was developed through collaboration between the civil engineering faculty and the university’s English language department.

Test Construct

The civil engineering faculty explicitly requested a scoring rubric that could be used to evaluate student performance during a 10-minute interview test-task. The goal was not, however, to assess English language skills alone but rather to assess students’ ability to survive and thrive in the United States as international students. The scoring

Carpenter & Matsugu: *Assessing Intercultural Understanding Through a Speaking Test*

rubric was designed to evaluate test-takers' language ability as part of their intercultural understanding. From the department's perspective, this meant that language skills should be evaluated, in principle, based on how well test-takers can make themselves understood using English. Therefore a greater emphasis was placed on students' expressiveness and communicative ability than on grammatical or lexical accuracy and complexity. In addition, although some of the factors listed in Sercu (2010; e.g., intercultural sensitivity, mindfulness, and attention to intercultural issues) served as a starting point when developing this rubric, evaluating these factors would likely have required students to answer questions about their degree of intercultural sensitivity, for example. The test development team preferred criteria that evaluated test-takers' observable behaviors that could theoretically be ascribed to intercultural competence. The civil engineering faculty were particularly concerned that their students develop the social skills necessary to negotiate aggressively and communicate actively with the international community. For this reason, the team decided to expand on one of the factors in Sercu—willingness to engage—and to place a greater emphasis on how the test-takers comported themselves during the 10-minute interview. This meant that rather than trying to illicit answers about how comfortable the student felt in intercultural situations, the examiners attempted to structure the interview so that it reasonably approximated an intercultural situation. The civil engineering faculty emphasized that they wanted to test students' ability to communicate under the pressure of a new, intercultural situation. The test development team operationalized this request by training raters to use, to the extent possible, only their most natural, fluent speech during the test task. In this way, students' ability to respond positively in this kind of interview task and to display a certain level of comfort and confidence throughout the interview came to represent one level of their intercultural understanding. Based on these considerations, the test development team developed a construct for intercultural competence consisting of three parts:

1. the ability to use English intelligibly;
2. the ability to participate actively in conversation through
 - a. providing detailed answers and
 - b. engaging actively with the examiners; and
3. the ability to navigate intercultural situations competently by
 - a. demonstrating attitudes such as confidence, directness, and clarity, and
 - b. appearing relaxed and comfortable throughout the test-task.

An analytic rubric with five bands was eventually developed to assess this construct definition (see Appendix). The first band assessed English ability—specifically, the degree of listener effort needed to understand the test-takers' speech. The second and third bands assessed the test-takers' active participation. Specifically, the second band assessed degree of detail in the test-takers' replies, and the third band assessed the extent to which test-takers actively engaged the examiners and attempted to participate in the conversation. The fourth and fifth bands assessed intercultural understanding explicitly. Specifically, the fourth band assessed each test-taker's "cultural fit," or the extent to which an individual test-taker's personality, as expressed in the interview task, seemed appropriate for the American cultural context. Finally, the fifth band assessed test-takers' apparent comfort level, as represented by their body language, facial expressions, and eye contact throughout the test.

Test Administration and Discussion

One Japanese member of the civil engineering faculty and two native-English-speaking faculty members administered the test to 20 students. In order to ensure that all three examiners interpreted the rubric in similar ways, everyone attended a training session the week before the test was administered. The purpose of this training session was both to normalize the raters' interpretation of the rubric and to establish the basic flow of the test administration. The examiners agreed on a set of simple, self-introduction questions (e.g., "Where are you from?" "What is your major?" "Why did you choose that major?") to begin each interview. The examiners were instructed to ask specific questions about why test-takers wanted to study abroad in the United States and what their long-term professional dreams were in the course of the test-task. The tests usually ended with some questions about students' expectations about studying in the United States. Figure 1 gives examples of some of the guiding questions agreed upon during the training session.

Why did you decide to study civil engineering?
 What are some of your professional goals?
 Why do you want to study abroad?
 How will study abroad help you achieve your goals?
 What do you imagine will be difficult for you while living overseas?
 What are you concerned about living overseas?
 What are you excited about?

Figure 1. Guiding interview questions.

Ultimately, however, no two interviews were the same in terms of the kinds of questions asked. This is primarily because one goal of this test was to assess students' ability to function well in an intercultural situation. In this sense, the quality of student participation (bands two and three), and their cultural fit and comfort level (bands four and five) were interrelated. In general, test-takers who actively engaged examiners and participated well in the flow of the conversation were also evaluated well for intercultural understanding. Test-takers who gave short or stilted answers generally scored lower on both their degree of participation and their intercultural understanding.

Certain spontaneous actions by test-takers were also evaluated positively across multiple categories. In one instance, a test-taker pushed his chair away from the table and crossed his legs as he spoke. In another, a test-taker approached the examiner's table to shake hands. One student, whose actual English fluency was relatively low and who was clearly nervous and uncomfortable shared with the examiners his desire to study abroad because "I cannot create my own identity in Japan, but in America I can create my own identity." This highlights a secondary point: that developing performance assessments centered around humanistic constructs like intercultural understanding allows for a more flexible interpretation of the criteria, thus giving a chance to test-takers to meet those criteria in their own way.

Lastly, the first band, that of English ability, was in many ways independent of the other four bands. As mentioned above, scores on this speaking test were meant to be aggregated with test-takers' TOEIC scores. Based on the test-development team's experience using this rubric so far, the actual relationship between TOEIC scores and performance on this speaking test is unclear. More specifically, it is unclear whether or not a higher TOEIC score predicts "intercultural understanding" as operationalized by

this speaking task. However, this is an empirical question beyond the scope of this paper.

The creation of this speaking test arose in response to the complex situation of simultaneously developing a study abroad program while reducing the program's overall cost. On one level, the test was arguably an attempt to create a winnowing device, meant to offset the financial burden of the civil engineering department, by identifying only the most capable students for studying in America. At the same time, the test was also critical and innovative. It was critical in the sense that its construction and implementation implicitly questioned the hegemony of the TOEIC as a measure of English proficiency at this university. It was innovative in the sense that it allowed students with lower TOEIC scores but strong social skills (i.e., intercultural understanding) to compete for a limited number of study abroad opportunities. It remains an empirical question, however, whether or not the incorporation of "intercultural understanding" into the construct of performance assessments will lead to enhanced experiences studying abroad. On another level, the final form of this test also reflects a tension between how the civil engineering faculty conceptualized the purpose of study abroad and how the English department conceptualized it. In general, the test development team was encouraged to design a test-task that advantaged students with particular social traits precisely because building relationships and developing strong social skills was the main purpose of the civil engineering department's study abroad experience. Although it is arguably the case that academic skills are more important for *studying* abroad, debating this point as part of the test development process inevitably led to debating what skills ultimately lead to professional success. In the end, the relationship between strong academic performance and professional success was well beyond the scope of this project.

Conclusion

Twenty test-takers are not enough to compute either the internal consistency of the bands on the rubric or to evaluate the relationship between speaking test scores and TOEIC scores statistically. Therefore, it is not possible to determine at this time whether or not this rubric is actually measuring a single construct in a consistent way. It is also not possible to determine how consistently individual raters were interpreting each of the criteria on the rubric. Test scores from a larger number of students and raters are necessary to evaluate both the validity and the reliability of this speaking test. Although the test-development team feels positive about the way the test has been implemented so far, research into the actual nature of the construct of intercultural understanding is necessary to improve this test or similar tests.

The test discussed in this paper is currently in its 3rd year of implementation. This means that the first group of students who were chosen to study abroad in the United States has recently returned to Japan. In order to validate assessment measures of this type, a qualitative investigation into these and future students' study abroad experiences is necessary.

If the systematic assessment of intercultural understanding becomes more widespread, longitudinal studies investigating the durability of students' intercultural understanding and how that understanding leads into future opportunities may also be important. Ultimately, the primary goal undergirding tests like this is to find and facilitate opportunities for talented students.

In this paper, the position is taken that identifying and cultivating talent is multidimensional. Teachers and administrators should become comfortable evaluating students' capabilities in a holistic way. This necessarily involves a certain degree of subjective judgment. The common perception is that standardized test scores are, to a certain extent, objective. Indeed, some colleagues at the university where this test was implemented were uncomfortable with the idea of evaluating a student's comfort level or their "cultural fit"—much less denying students the opportunity to study abroad based on such subjective impressions. Yet, these same teachers agreed that relying on TOEIC scores alone was likely not effective either. It is therefore argued here that the solution to this problem is not designing a better, standardized English language test. Instead, assessments—particularly performance assessments—should be designed that deliberately make use of teachers' own empathy, intuition, and maturity in order to evaluate students for major opportunities like studying abroad in a holistic way. Every student has the right to be fairly assessed, especially when future opportunities are tied to assessment results. One way to make performance assessment tests fairer may be by incorporating intercultural understanding into the test construct.

Bio Data

James Carpenter has an MA in TESL and an MEd in educational technology from Northern Arizona University. He has taught ESL in the U.S. as well as EFL in Japan. His research interests include project-based learning, educational technology, and alternative teaching methods. He is currently an instructor in Asia University.

Sawako Matsugu is an instructor in Rikkyo University. She has a PhD in language assessment from Northern Arizona University. She has taught EFL in Japan as well as ESL in the U.S. Her research interests include language assessment, in particular speaking assessment and effects of rater background on assessment.

References

- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. New York, NY: Oxford University Press.
- Brown, J. D., Hudson, T., Norris, J., & Bonk, J. W. (2002). *An investigation of second language task-based performance assessment*. Honolulu, HI: Second Language Teaching & Curriculum Center.
- Celce-Murcia, M., Brinton, B., & Snow, M. (2014). *Teaching English as a second or foreign language*. Boston, MA: National Geographic Learning.
- Liddicoat, A., & Scarino, A. (2010). Eliciting the intercultural in foreign language education at school. In A. Paran & L. Sercu (Eds.), *Testing the untestable in language education* (pp. 52-73). Bristol, England: Multilingual Matters. <https://doi.org/10.21832/9781847692672-006>
- Livermore, D. (2013). *Expand your borders: Discover 10 cultural clusters*. East Lansing, MI: Cultural Intelligence Center, LLC.
- MEXT. (2017). Measures based on the four basic policy directions. Retrieved from <http://www.mext.go.jp/en/policy/education/lawandplan/title01/detail01/sdetail01/1373805.htm>
- Sercu, L. (2010). Assessing intercultural competence: More questions than answers. In A. Paran & L. Sercu (Eds.), *Testing the untestable in language education* (pp. 17-34). Bristol, England: Multilingual Matters. <https://doi.org/10.21832/9781847692672-004>

Appendix

Test Rubric

	40 points	30 points	20 points	10 points	0 points
English ability	English is clear, and appropriate , and involves minimal listener effort .	English is mostly clear, and appropriate , but involves some listener effort .	English is often unclear , and requires more listener effort .	English is very difficult to understand , and requires a lot of listener effort .	It is very difficult to understand the student because of their language use .
Degree of detail	Answers are thorough, with details and examples .	Answers are mostly thorough , but few details or examples .	Answers are limited , with only one detail or example .	Answers are short , with no details or examples .	Answers do not address the question .
Participation	Student attempts to participate fully in the flow of conversation, and actively engages the examiners .	Student sometimes cannot participate in the flow of conversation, but still tries to actively engage the examiners .	Student often cannot participate in the flow of conversation, and does not actively engage the examiners .	Student only answers the questions , and does not participate in the flow of conversation.	Student does not engage in meaningful conversation.
Cultural fit	Communication style is appropriate for living in the US.	Communication style is usually appropriate for living in the US, although sometimes inappropriate .	Communication style is often not appropriate for living in the US, but overall interaction is still positive .	Communication style is often not appropriate for living in the US, and the overall interaction is negative .	Communication style is not appropriate for living in the US.
Comfort	Comfortable <ul style="list-style-type: none"> • Eye contact • Facial expression • Body language 	Mostly comfortable	Sometimes uncomfortable , and some anxiety .	Often uncomfortable communicating and much anxiety .	Too nervous to communicate.