UCLA
Issues in Applied Linguistics

Title
Language Assessment as a System: Best Practices, Stakeholders, Models, and Testimonials

Authors
Avineri, Netta
Londe, Zsuzsa
Hardacre, Bahiyyih
et al.

Peer reviewed

PANEL DISCUSSION

Language Assessment as a System:
Best Practices, Stakeholders, Models, and Testimonials

Introduction

Netta Avineri, Panel Moderator
University of California, Los Angeles

Throughout the controversy surrounding the language proficiency of Arizona's English teachers the question of assessment consistently arose. Many people wondered how the Arizona English teachers' proficiency was being assessed, what standards were being used, how often these assessments occurred, and who was doing the assessing. The conference organizers therefore felt strongly that a panel focused on the complex issues related to language assessment was an essential part of the broader conversation. The panel, "Assessing and Addressing English Proficiency" included the following participants: Dr. Zsuzsa Cziraky Londe (International Teaching Assistant Testing Coordinator at the University of California, Los Angeles), Dr. Youngsoon So (former UCLA doctoral student specializing in language assessment), Lauren Mason Carris (current UCLA doctoral candidate and former Test of Oral Proficiency coordinator), Bahiyyih Hardacre (current UCLA doctoral student and current Test of Oral Proficiency coordinator), and Mostafa Majidpour (current UCLA electrical engineering doctoral student and former Test of Oral Proficiency test taker), with Netta Avineri (current UCLA doctoral candidate and former Test of Oral Proficiency coordinator) serving as the moderator.

Over the course of our discussion during the panel we realized the multiple ways that language assessment is not a tool but a system, one that must include a deep consideration of the context of language use and education and key stakeholders. In this article the panel members provide a diversity of perspectives of issues related to language assessment through their discussion of stakeholders (Zsuzsa Cziraky Londe), a detailed examination of one model for language assessment (the Test of Oral Proficiency at UCLA, by Lauren Mason Carris and Bahiyyih Hardacre), and testimonials from both test raters and test takers (Youngsoon So and Mostafa Majidpour). When one considers the incredibly diverse population of a city like Los Angeles and its ability to attract international teachers and students from all over the globe, it is quite fitting that the conference's discussion about the complexities of assessment occurred here.

There are a number of language assessment *best practices* that we have identified over the course of our discussions, all of which are centrally relevant to administrators, practitioners, test takers, and other stakeholders. These are based in part on Bachman (2005) and Bachman and Palmer's (2010) model of an Assessment Use Argument, which "provides the rationale and justification for the decisions we make in designing and developing the test" (Bachman, 2009, p. 4). Bachman and Palmer encourage assessment developers to consider the consequences, decisions, interpretations, assessment reports/scores, and assessment performance both at the stages of test design/development and test interpretation and use. In addition to the central components considered above, we highlight two other central issues. The first is the importance of understanding the larger context in which the assessment is being administered. In this consideration of context, we include the classroom, school, neighborhood, region, and national context, as these are all quite relevant to the assessment experience. Therefore, in the case of Arizona for example, it would be essential to consider the political and cultural complexities of each of these contexts, in addition to the ways that the raters and the test takers may experience the language assessment itself. The second issue to consider is that of the test takers' and the raters' subjectivities. The two testimonials included in this article therefore emphasize the lived experience of the assessment.

The five *best practices* we have identified are as follows: 1. Take into account ALL of the various stakeholders for any assessment situation. These can include administrators, teachers, test takers, students, parents, and others. The issue of stakeholders for international students' language assessments is considered in detail in Londe's section of this article. 2. Provide a number of resources beyond the assessment tool itself. As Hardacre and Mason Carris describe, for the UCLA Test of Oral Proficiency (TOP) the assessment tool is complemented by oral skills courses and individual counseling sessions with TOP coordinators. Resources like these provide a wide range of services for both test takers and administrators alike. 3. It is essential that assessors are well-trained in issues of language assessment in general and the specific issues related to a given language assessment tool in particular. This provides reliability and validity for the assessment tool, which is critically important for all of the stakeholders. 4. A collaborative team working together on the assessment system is important. The team, made up of practitioners, teachers, researchers, and others, can draw upon a large body of knowledge of both institutional practice and scientific research (as opposed to simply a manual that test administrators follow). 5. We highly encourage ongoing accountability, reflective practice, dialogue, and coordination among all the parts of the assessment system. It is only through ongoing discussions with test takers, raters, administrators, and others that an assessment tool can be fine-tuned for the populations it is intended to serve.

This article seeks to present a unique perspective on assessment systems, as it provides a forum for the voices and experiences of administrators, raters, and test takers to be heard.

Stakeholders In International Teaching Assistants' Oral Language Exams

Zsuzsa Cziraky Londe, Panelist
University of Southern California

Introduction

During this conference issues of discrimination against people with accents in the "teaching business" was the subject matter of several presentations. Our panel discussion specifically addressed the issues of the international teaching assistants' (ITAs') oral exam; whether the exam is necessary, and how the exam is administered.

The 15 to 20-minute period, during which the test taker provides a sample of her/his English ability, typically becomes highly charged, but has no discriminatory qualities. In fact, it is the high-stakes character of the test that heightens the test administrators' awareness of the need to design reliable assessment instruments, and to administer and evaluate them professionally and objectively.

International teaching assistants' English language assessment is not an instrument to discriminate, but a tool that allows each stakeholder to understand, evaluate and appreciate their own roles and responsibilities in this process. Over six hundred seventy thousand international students attended US institutions in 2008-2009, which was an 8% increase from the previous academic year (Hvisten-dahl, 2009). With globalization, "less borders" that limit educational institutions around the world, this number will most likely grow. he inexorable movement toward globalizationis lready increasing this number. It is therefore important that we understand who the stakeholders are in the ITA test, and why is it considered a high stake test.

As ITA-testing coordinator at the University of Southern California (USC), with my training in second language assessment, I oversee the design and administration of ITA oral exams. These exams serve to evaluate whether the ITA's English proficiency is appropriate to teach undergraduate and/or graduate students at a university. The importance of monitoring English levels became nation-wide news in 2005, when a Berkeley student initiated a lawsuit against the university, because she did not understand the TA in her major class and had to switch disciplines, thereby prolonging her study years. Most recently, a new state initiative in Arizona proposed not to hire TAs with accents (regardless of language proficiency), which could directly affect the lives of international TAs. These two cases drammatically illustrate why it is important to have an assessment instrument, and why it is essential to identify the entities whose policies, practices, and lives could be affected by such an instrument. The entities involved are the stakeholders; "they hold a 'stake' in the use of a given test in any particular situation" (Bachman, 1996). Understanding their interests and what impact the test has on their lives in this process sheds light on the complexity and importance of the exam and the issues surrounding it.

Stakeholders in International Teaching Assistants' Oral Language Exams

The following is a list of the populations affected by the live, performance-based oral assessment: 1. international graduate students, interested in serving as teaching assistants, 2. undergraduate students taught by international teaching assistants (ITA's), 3. the department, in which the course is taught, 4. the university, 5. the ITA exam administrators and 6. society and the country,

1. International graduate students/ International Teaching Assistants: The ITAs themselves have interests in teaching/TA-ing a class to various degrees and for several reasons: a) it is an honor to be appointed by their advising professor, b) it is an important learning experience (teaching is the best way of learning), c) it usually provides a significant and sometimes their only financial support (this affects the TA's families are also affected), d) the TA-ship is often a departmental requirement, and e) many departments need knowledgeable instructors for lower level survey courses.

Accordingly, the ITA's stake in the ITA exam is of varying significance. It is the department and a professor who makes the initial evaluation whether a student is prepared to be a TA; i.e. knows the material to teach. It is the university's responsibility to test whether or not the student can properly convey this knowledge in English. USC, for example, summarizes its role in their catalogue as follows: "The ability to communicate effectively in English—to read, write and speak the language fluently—is vital to your success as a graduate student. Receipt of a USC graduate or professional degree signifies that its holder is fully qualified to conduct academic and professional pursuits in English."[1] The oral exam is important to the ITA.

The TOEFL (Test of English as a Foreign Language) exam is one of the indicators whether a student's English language proficiency is adequate and is only used for admissions purposes. In many cases the universities make use of a "local" ESL test to uphold their own standards to assure that students can follow the lectures, complete written homework, and be competent speakers. The goal is a "fluid" classroom in which the student does not have to be stalled by the language of communication but can reflect on the material at hand. It is the university's role to uphold the teaching-language standards, it is the ITA's role to live up to these standards, and it is the testing unit's role to interpret the university's standards to practical and useable grades and evaluate whether the student lived up to these standards. The potential ITAs are important stakeholders in the high-stakes oral exam.

2. Often undergraduate students in courses taught by ITA's could have a hard time understanding non-native speaker teaching assistants, and have become frustrated enough to create a problem, reaching departmental or university levels. There is a point at which "cognitive fluency" (Lev-Ari and Keysar, 2010) is hindered by an "unusual" language and the brain has more difficulty processing the information. It is therefore understandable that students, who have to struggle with difficult concepts they could not grasp during the professor's lecture, would become frustrated trying to understand the TA's explanation. These students,

therefore, are greatly impacted by the TA-choice for their classes - their grades and learning outcome could be directly affected. They are directly impacted and are stakeholders in this process.

3. Departments have the responsibility to assign the right TA to a class. The department is also making the decision to whom grant financial support (TA-salary), which at times determines whether a student can continue his or her studies. When it comes to international TAs, it is the department's responsibility to assure that the TA's language proficiency is satisfactory. The departments struggle between the obligation of finding well prepared TAs for their classes, the financial commitment towards the international student, the requirement of satisfying university language standards, and in the case of USC, for example, the responsibility to pay for the ITA's language classes if additional ESL classes are necessary. The departments, therefore, have a substantial stake in the outcome of the ITA exam. If the ITA does not perform well, the department has to find another TA to teach the class, has to find other ways to finance the student who failed the test, and has to pay for the student's ESL classes to prepare for possible future TA-ships.

4. The university's role in this process is to uphold standards of teaching. The university, in collaboration with the individual departments, has to ensure that undergraduate and graduate students receive instruction from teachers with adequate English proficiency for the overall purpose of the universities' mission to advance knowledge. On the other hand students require that such standards be upheld, because they invest in the university and in their own future by being at a particular institution. The university has to find the balance between the obligation of language standards, i.e. the teaching language, the university's prestige, i.e. each matriculated student represents the institution, and the lucrative business of admitting foreign students to the university (an $18 billion contribution to the US economy, see Goodman, 2010). The university is therefore a stakeholder in the ITA oral exam, because through the exam it upholds university standards and complies with the demands of the students for quality education.

5. ITA exam administrators have a responsible role in the process of conducting the TA English oral exams but they are not stakeholders, they do not have an interest in the process of evaluation. The stakes are high for all the stakeholders, therefore their roles are highly responsible, but the administrators/language experts' role is that of integrity, to remain unaffected by the stakeholders' interests. It is sometimes difficult not to think about the immediate consequences of a student's low scores, or not to be affected by departments' disbelief of students' performance. However, the administrators/testers have to remain professional executors of the plan set forth by all the stakeholders. Exam raters are also key stakeholders, for their reliability, training and subjectivities are critical elements in test takers' ultimate scores.

6. Society and the country are also be affected by the ITA exam decisions and are stakeholders in this process. If a student does not pass the exam and is not cleared to TA, he or she may have to leave the country for lack of funds. Or,

if the universities would simply not admit international students to make things easier, society and the country could lose a potential leading scientist or even a future "Nobel Laureate." International students are essential to the advancement of knowledge, not only because of the individual "brain power" they bring to their field but also because of the diversity necessary for academic and scientific advancement. Indirectly, therefore, society and the US are also affected by the ITA oral exam and are indirect stakeholders in the process.

Conclusion

Because of the stakeholders' interest in the test, and because of the wide variety of consequences mentioned above, the ITA exam is considered a high-stakes exam. While the procedure directly impacts some stakeholders and indirectly others, it is important to be aware and considerate of all the consequences at every stage of this process, to make it fair, effective, and non-discriminatory.

The UCLA Test of Oral Proficiency: A Model for Assessing and Addressing English Proficiency of International Teaching Assistants

Bahiyyih Hardacre and Lauren Mason Carris, Panelists
University of California, Los Angeles

Background

In this section of the article, we would like to discuss how the University of California, Los Angeles (UCLA) has addressed the need to evaluate the English speaking proficiency of international students who are prospective teaching assistants (TAs). For the past 6 years, UCLA has developed and administered its own oral proficiency test, designed to serve the specific purpose of assessing international students' level of English proficiency to perform teaching assistant duties. Our hope is that the actual practices of this university can provide concrete examples for how the English proficiency of this specific population can be approached in the US university system.

The Test of Oral Proficiency (TOP) at UCLA

Before the creation of this assessment tool, there was a general concern within this institution, and particularly from the Graduate Council, to address undergraduate students' claims that incomprehensible teaching assistants leading labs, discussions, and lectures prevented them from succeeding in class. Aimed at addressing this concern, the UCLA Office of Instructional Development (OID) decided to look for alternatives to the then widely known and used ETS-created *Speaking Proficiency English Assessment Kit* (a.k.a. the SPEAK Test), for it was apparent that it was not successfully informing which international graduate students were ready to serve as TAs.

In order to resolve the problem, Tim Farnsworth (see Farnsworth, 2004) proposed a pilot study in which he defined the construct of the kind of oral test this institution needed, and provided initial backing for warrants about the meaningfulness of a UCLA-specific test of oral proficiency. This pilot study was then approved by the UCLA Graduate Council, implemented by OID, and made mandatory for all international graduate students applying for TAships on this campus.

Up to this date, all international graduate students who plan to work as teaching assistants for any department at UCLA are required to pass the *Test of Oral Proficiency* (TOP). One important feature of this test is that it was designed to only assess relevant oral language use, leaving aside content knowledge. More specifically, the TOP assesses language systems (e.g., pronunciation, grammar, and vocabulary), rhetorical organization and question handling. Relevant domains of language use include teaching discussion sections or labs, holding office hours, and engaging in other interactions with students regarding course content and administrative details.

The test consists of 3 tasks: a brief self-introduction, a syllabus presentation, and a prepared 10-minute mini-lesson presentation on a basic topic in the test-taker's field. Only the latter two tasks are scored, each weighing 50% of the total score. Farnsworth's (2004) pilot study concluded that an analytic scoring rubric would make the necessary distinction between language proficiency and teaching proficiency more explicitly than a holistic rubric. Thus, each scored task is rated (0-4) on four separate subscales (pronunciation, lexicon/grammar, rhetorical organization, and question handling). Analytic scores are weighted and then transformed into scaled scores.

Scaled scores are used to determine whether an international graduate student has adequate oral proficiency to successfully perform typical TA duties. A test taker's scaled score falls into one of three categories based on cut-scores: Non-passing, Marginal Pass, or Pass. Scores are sent directly to the test takers and their appointed departments by email. Departments are directed not to appoint TAships to international graduate students that have not taken or have not passed the TOP exam. However, it is not the responsibility of the test coordinators to verify if this requirement is being enforced; UCLA Graduate Division is responsible for this verification process.

In order to provide opportunities for learning and improvement, and not discriminate against any particular linguistic backgrounds, if test takers earn a "Non-passing" score they are allowed to retake the test in the following administrative quarter. In addition, if a test taker's scaled score falls in the range designated "Provisional Pass," the test-taker is required to complete ESL coursework before or during his/her first quarter as a TA. During counseling sessions, TOP administrators recommend particular coursework based upon the test taker's performance. They choose from three oral skills classes offered every year (ESL 38A, 38B, and 39B[1]) designed for this population. ESL 38A, "Pronunciation: Stress and Intonation in English", provides training on suprasegmental features of pronunciation;

ESL 38B, "Pronunciation: Sound System of English", provides further training on segmental features of pronunciation, such as vowel or consonant clusters sounds, among others; and ESL 39B, "Communication Strategies for International Teaching Assistants", provides further training on classroom teaching skills (e.g., giving presentations, using classroom language, or engaging in classroom management).

Thus, test scores are also used to direct further study, learning, and improvement. The TOP coordinators are also able to recommend various forms of language improvement strategies, depending on each student's needs. Therefore, the exam screens for language-specific skills, in addition to serving as an opportunity for feedback and a personal oral skills improvement tool for all test takers who complete it.

Beyond Assessment – Providing Resources for Test Takers
As previously mentioned in this article, the Test of Oral Proficiency resides in the Office of Instructional Development, a unit within the university committed to enhancing teaching and learning opportunities on a campus-wide scale. Thus, the TOP is embedded within a larger system of accountability, transparency, and support for international graduate students interested in becoming teaching assistants. The test coordinators are well informed about the free resources that the campus has to offer its graduate students and are able to counsel test takers as to what to do next. Over the course of the six years that the test has been implemented, the test coordinators and test takers have all benefited from the campus' expertise in language testing, language training, and language teaching in the department of applied linguistics.

The direct link to applied linguistics, and former academic tether, Teaching English as a Second Language program, provides a unique opportunity for the broader campus to benefit from cutting-edge English language pedagogy, specifically designed for International Teaching Assistants (ITAs). Janet Goodwin, now a lecturer in Writing Programs, has dedicated a lifetime of pedagogical research, program development, and instructional training to the unique population of ITAs. Together with her colleagues, she has developed a strong, supportive segment of English language support within the Academic English as a Second Language program, which addresses the unique needs of ITAs and offers a range of oral skills courses. These courses are designed specifically for interaction and communication in university classroom settings and address topics like: pronunciation, communication strategies for the classroom, interactive teaching, discussion leading, presentation skills, giving instructions, making small talk, explaining a process, presenting with visual aids, and holding office hours.

Importantly, many of these skills have a significant impact on perceptions of "language proficiency" by native English-speaking undergraduate students. If an undergraduate student meets in office hours with his/her international TA or raises a concern during discussion section, it is important for that TA to be able to communicate effectively – a task that often encompasses cultural knowledge of the

language and effective communicative strategies beyond research-level knowledge of the subject matter. Furthermore, in many cases, the burden of clarification, demonstration, and lab instruction often falls on teaching assistants, who are under time constraints to bring students up to speed for the pace of lectures. By learning discursive strategies used by current teaching assistants in their respective fields and across the disciplines, international graduate students who take the Oral Skills courses are equipped with a set of linguistic and communicative tools that positively impact how they are understood.

What is particularly innovative in these classes is the use of technology and the special attention given to "real life language". In these courses, instructors and students regularly make use of audio-recorded and videotaped language samples to provide models of native-speaker intonation, pace and delivery, word choice, embodied speech, and other cultural cues that impact effective communication. These resources are available to students outside of class as well. Similarly, audio and video recordings of instructor language, and student "performances" of instructor language help students learn to effectively self-assess their developing skills. Thus, these courses provide a wealth of resources for effective communication in the classroom, for international graduate students who are preparing for the TOP, for those required to take supplemental oral skills courses due to their "Marginal Pass" TOP score, and for those simply interested in new approaches to communication for the university classroom. Additionally, priority enrollment is provided to ITAs who have received a "Marginal Pass" on the TOP. Thus, ITAs whose test score reflects the need for additional resources but who may have a pending TAship or even be in the classroom concurrently, have the opportunity to receive explicit instruction and feedback on their oral skills as they pertain to the classroom.

Maintaining Test Reliability: TOP Rater Training Program and Questioners' Comprehensibility Survey

One of the many challenges a non-computerized test faces is the use of human raters, for many are the factors that could alter a score given to a particular test performance. For instance, raters might be familiar with a test taker's accent and therefore have the impression the test taker's proficiency was very comprehensible and therefore acceptable, whereas other raters who are not as familiar with that particular accent might struggle to understand what the test taker is saying; this might result on the rater giving the same test taker a lower score. This is evidence that one of the great challenges this exam must deal with is inter-rater reliability and agreement (Bachman & Palmer, 1996). According to Bachman (1990), raters may differ in the interpretation and/or application of the scoring criteria: they might differ in the degree in which they comply with the scoring rubric, in the way they interpret the criteria employed in operational scoring sessions, in the degree of severity or leniency exhibited when scoring test takers' performance, or in the understanding and use of rating scale categories.

Another equally challenging characteristic of involving human raters as opposed to relying solely on fully computerized proficiency exams is intra-rater reliability. The same rater might give different scores to the same test taker depending on the time of the day, how tired he or she is, which will affect how much effort was put into following the test taker's spoken delivery. However, both inter-rater and intra-rater reliability are most likely to be variable if a score relies entirely on raters' impression of the test takers' comprehensibility. For this reason, we must make sure all raters are constantly normed with the test rubric, scoring categories, and test decisions.

The rubric specifically created for this test describes a test taker's range of abilities on each of the test categories: pronunciation, lexicon/grammar, rhetorical organization, and question handling. The rubric provides detailed descriptions of each category across the four assignable scores, from 1 to 4 (lower to higher proficiency, respectively). It is imperative that all raters become very familiar with the rubric in order to maintain scores fairness and reliability, besides offering a sound justification of the score given to a test taker. Raters are required to keep a copy of the rubric in front of them at all times, along with the scoring sheets, for the entire duration of the test, and training and norming sessions as held quarterly in order to help such raters to understand and assimilate what each score category entails, as dictated by the rubric.

Thus, such training and norming sessions are crucial to maintaining raters' inter- and intra-reliability. Although rater training cannot eliminate rater bias, it can make raters more self-consistent. Studies have shown that results of rater training may not endure for long after a training session, so the practice of holding a norming session before each test administration is crucial to allow raters to re-establish an internalized set of criteria for their ratings (Lumley & McNamara, 1995). As part of the TOP rater training program, raters must take part in one annual training that takes place in the Fall, as well as quarterly norming sessions that are held online. The Fall training sessions take place on two different days and are comprised of 3 distinct sections, starting with a discussion of the scores and categories on the rubric, along with examples from previous exams. During the second section of this training program, raters attend workshops on how to describe and document typical issues found on each of the test categories (pronunciation, lexicon/grammar, rhetorical organization, question handling), also paired with clips of previous tests. What follows is a final section with a series of norming videos, when training raters are asked to discuss the scores they would give to the formerly videotaped test takers, and how they would justify such scores. This discussion helps raters identify nuances of oral performance and how to relate what was observed and documented back to the rubric.

Quarterly norming sessions are held online one week prior to the test administration. Raters are given access to a password-protected training site, where they find the rubric, scoring sheets, and 3 videos to watch and score; they are required to send their scores to the TOP coordinators, along with the required annotation

justifying the scores given, tying them back to the rubric. The TOP coordinators take a look at the scores and documentation, and send these raters some feedback, suggesting areas that need improving, as well as how close they were to the ideal scores and descriptive language. Raters that seem to be successfully normed are invited to work during the following test administration, whereas raters who were too far from the appropriate decision category or did a poor job justifying their scores are not invited to rate during that test administration, but they will be given a posterior chance to be normed again.

Finally, TOP coordinators have one additional resource to ensure score decision reliability: the questioners' "comprehensibility survey". The questioners that work with the TOP are 2nd and 3rd year undergraduate students at UCLA, trained to perform consistently across tests. Two questioners pass as "the class" during the test, asking questions and interacting with a test taker pretending to be regular undergraduate students at UCLA. For this reason, questioners have valuable insight into any given exam. At the end of the test, they are asked to fill out a brief comprehensibility survey, a resource coordinators use to examine raters' scoring reliability. These questioners are asked to report on how comprehensible they think each test taker was, and if they think that test taker has the linguistic resources to carry out TA duties. The TOP coordinators have just begun implementing this survey, having started to collect them in November 2010. Thus far, and based on very preliminary data, questioners agree with each other 70% of the time, and agree with the raters' category decision (pass, marginal pass, not passing) 70% of the time. We are still to collect more data in order to understand what this correlation means and to make appropriate adjustments to both our questioner and rater training programs.

Additional Research

In 2010, the exam coordinators, aiming at providing a useful document for the sustainability and backing of this test and also to identify areas for its improvement, created an *Assessment Use Argument* (Bachman & Palmer, 2010) for the TOP. An assessment use argument (AUA) sheds light into how a test is beneficial to its stakeholders, how the decision categories are to be used, and how such decisions are meaningful, generalizable, relevant and sufficient for what the test proposes to assess. Along with the claims it makes, this AUA describes what could be such claims' warrants, and addresses hypothetical rebuttals to these claims.

While a number of studies will be necessary to provide additional backing for the TOP, this AUA may help prioritize a research agenda. For example, a well-designed survey of test takers may provide backing for many warrants which may significantly strengthen the TOP's assessment use argument. Instituting a periodic procedural and bias review may be time-consuming for TOP coordinators, but if well-designed it would provide backing for an additional warrants across two of the test's AUA claims. Thus, the development of an AUA has clearly identified needed areas of validation research for the TOP and provided a means to prioritize research.

A Test Taker and Rater's Testimonial

Youngsoon So, Panelist
Educational Testing Service

I would like to share my experiences with the TOP with related to the two roles I have taken on: a test taker and a rater. In addition, I also want to discuss value-added qualities of the TOP, seen as an applied linguist specializing in language assessment, compared with other previous speaking tests that were used – or are still used in other universities – to make decisions about prospective ITAs' readiness to teach an English-medium class. Finally, I want to share an episode from my own teaching experience and discuss what implications it provides for having a non-native teacher in a US classroom.

First, I took the TOP in the first year I came to UCLA, as any other international graduate student did who wants to be hired as a TA. My first reaction to the test was that it was something that caused extra burdens and stress for students who had to prove their English proficiency in order to apply for the university. However, in the end, I got to realize that the test was different from other tests that I had taken before and that it was designed to serve a very specific purpose. I found myself getting more prepared for teaching undergraduate students in the process of preparing for the test.

Second, after I passed the test and was hired as a TA for several quarters, I started working as a trained rater for the test and have had a chance to see the test from a different point of view. More specifically, I learned how the issue of a 'foreign accent' is taken care of in scoring test takers' performances and why the two scoring criteria, rhetorical organization and question handing, were included in the scoring rubrics even though they are not commonly considered in other speaking tests.

When these personal experiences about the TOP are related to theories in language assessment, I would like to highlight one very strong point of the test: high authenticity. The term 'authenticity' means the degree to which language testing contexts are similar to the real-world or non-testing situations in which test takers are expected to perform in the target language. This quality of test design is related to the degree of confidence that a language test could have in generalizing a test taker's performance in a language test to his/her real-world language use. Therefore, authenticity is eventually related to the validity of a decision about a test taker's language ability to perform in a specific context on the basis of his/her performance on the test.

From the experiences of both taking the TOP and being involved in the administration of the test, I noticed that my teaching, or my ability to teach, has been improved and I feel less nervous standing in front of many students and teaching them difficult concepts in English. However, it does not mean that I have been

welcomed by all students in the various classes I have taught. The harshest comment I received came from a course titled 'language testing' targeted to prospective English as a Second language (ESL) teachers in a TESOL Master's program. The comment is copied below:

> I feel very cheated by the department for choosing someone with her lack of teaching ability to teach complex material. How can a person be expected to learn from someone who has a drastic accent?

When I first read the comment, I was just upset about this terribly low evaluation and felt very embarrassed. At the first thought, she gave to me. But at the second thought, I got sympathetic about future students who would be taught English by such a teacher who is not tolerant to 'accented' English. The reason I included an embarrassing episode of mine is because it provides implications for having non-native teachers in the US school system. I believe one of the important objectives of education, if not the most important, is to help students be aware of cultural and linguistic diversity in their communities and in the today's world. Non-native teachers have a clear edge on raising students' cultural awareness about and positive attitude toward languages and cultures other than their own.

A Test Taker's Testimonial

Mostafa Majidpour, Panelist
University of California, Los Angeles

The first day I knew I am going to be a TA at the school, I was kind of nervous about what will happen in the class. I had the experience of not being understood by faculty members, international and domestic students (and other people who did not belong to the same country as I did) because of my accent and pronunciation. Although I was trying to imitate the native accent, it seemed it is not going to be easy.

Fortunately, I was exposed to the TOP exam here at UCLA, where I was expected to show my ability to communicate with students comprehensively. The test consisted of teaching some elementary topics for a few minutes followed by students asking their questions. Each test taker was rated based on four skills: pronunciation, lexicon/grammar, rhetorical organization, and question handling.

One of the advantages of this test is that I could consult with an expert. We would discuss my strengths and weaknesses, and how to improve by taking effective courses. I was recommended to take one ESL course. Taking this class was a great opportunity, and I met a wonderful teacher who devoted her time to teach us many different necessary materials for the "classroom." By classroom, I mean we not only worked on different accent and language communication issues, but we also had the blessing to learn more about classroom management (such as how to react to some situation in class, or how to handle different session like Q/A, etc.).

After being a TA for two quarters, I found that the scenarios of the TOP exam, as well as some of the situations covered in the ESL courses' were very close to what happens in the class. Hence, I believe the exam and courses are extremely helpful in diagnosing communication flaws and resolving them. I encountered the same (potential) difficulties in the class that I had faced in the exam, and if there was no ESL courses, students as well as myself would be suffering from a not-completely-comprehensive class.

Notes

1. http://www.usc.edu/admission/graduate/international/english.html

References

Alderson, J. C., Clapham, C. and Wall, D. (1995). *Language test construction and evalu-ation*. Cambridge: Cambridge University Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press: 177-181.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34.

Bachman, L. F. (2009). Justifying the use of language assessments. Presentation at Guangzhou University. April 2009.

Bachman, L. F. and Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L. F. and Palmer, A.S. (2010). *Language assessment in practice: Develop-ing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Farnsworth, T. (2004). *The effect of teaching skills on holistic ratings of language ability in performance tests for international teaching assistant selection* (Masters thesis). University of California, Los Angeles.

Goodman, A. E. (2010). Why foreign study matters. *Parade Magazine*. April 18, 2010.

Hvistendahl, M. (2009). Number of foreign students in U.S. hit a new high last year. *The Chronicle*. November 16, 2009.

Lev-Ari, S. and Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology* 40: 1093-1096

Lumley, T. & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, 12/1: 54–71.

Netta Avineri is a doctoral candidate in applied linguistics at UCLA. Her dissertation research focuses on contemporary secular Yiddish pedagogy, centering on how endangered languages are taught and learned in intergenerational contexts.

Lauren Mason Carris is a doctoral candidate in applied linguistics at UCLA, specializing in relationships between language, race and ethnicity. Her dissertation integrates critical discourse analysis with multi-sited ethnographic research, to provide insight into the multiple semiotic resources used by Chicana/o Teatro Artists to fashion identities and reinforce, negotiate, and/or subvert linguistic and cultural norms.

Bahiyyih Hardacre is a doctoral student in the Department of Applied Linguistics at UCLA, and coordinator of the UCLA Office of Instructional Development's Test of Oral Proficiency (TOP) program. Her main research interests include language assessment and linguistic neurophysiology, with emphasis on learning through encultured development and social engagement behaviors.

Zsuzsa Londe is lecturer and International Teaching Assistant testing coordinator at USC. Her research and publications are in the areas of generative linguistics, psycholinguistics, and language assessment.

Mostafa Majidpour is a second-year PhD student in electrical engineering at UCLA, who has conducted research projects in pattern recognition and machine learning, as well as control engineering. His current research interests include the application of these ideas in Smart Grid-related topics.

Youngsoon So graduated with a PhD in language assessment from the Department of Applied Linguistics at UCLA. In addition, as a language assessor and applied linguist, she has been involved in the TOP administration as a trained rater for three years.