

INTERACTIONAL/ABILITY AUTHENTICITY AND THE DEVELOPMENT OF A CRITERION-REFERENCED PERFORMANCE TEST FOR EVALUATING THE ORAL PROFICIENCY OF TEACHING ASSISTANTS

Shahrzad Saif

Department of Linguistics
University of Victoria

1. Introduction

In the field of language testing, the assessment of oral production ability has always been one of the most important aspects. However, many well-established tests either do not have an oral component or if they do, not enough weight is allocated to the oral aspect compared with other components tested in the written mode. This is mainly because--disregarding the elementary level of mimicry and repetition--it is extremely difficult to construct practical, reliable, and valid communicative tests of speaking.

One problem is that questions related to the criteria for measuring the speaking skills and to the weight given to such components as correct pronunciation and sound recognition are still unanswered. It is possible for people to produce practically all the correct sounds but still be unable to communicate their intended meaning effectively. On the other hand, people can clearly express themselves while having numerous errors in both phonology and syntax. Furthermore, as Heaton (1990) states, success in communication depends as much on the listener as on the speaker, a particular listener may have a better ability to decode the foreign speaker's message or may share common ideas with him/her thus making communication simpler.

Many tests of oral production are also fraught with the problem of interrelationship between speaking and listening skills. In normal speech situations, it is impossible to hold a meaningful conversation without understanding what is being said and without making oneself understood at the same time. This interdependency increases the difficulty to analyze precisely what is being tested at any one time. Moreover, due to the transient nature of spoken language, the examiner in an oral test will be under great pressure of making subjective judgments as quickly as possible. Even if samples of speech are recorded during a test, the tape recording itself, apart from the unrealistic situation that it would create for the elicitation of speech, cannot recapture the full context of the actual situation essential to any assessment of the communication that takes place.

Still another difficulty in testing speaking is that of administration. It is almost impossible to test and score large numbers of students during a limited time.

However, particular merits or limitations of different testing instruments cannot be fully appreciated apart from the uses to which they are put. One or the other of the above restrictions can be relaxed depending on the objectives of a particular test, its theoretical basis, the population being tested, and the type of the test (e.g. whether it is an achievement or a placement test). As such no one can pass judgments on a test without having particular group(s) of examinees with definite objectives in mind. A test which is considered as practical for testing a specific population with respect to a particular objective may prove impracticable for testing another group of students with different objectives.

This paper addresses the problem of appropriately assessing the oral proficiency of non-native graduate students working as teaching assistants (TAs) in English-speaking environments. Working from the theoretical standard of "authenticity" for language testing (Bachman, 1990) and the studies so far conducted with the purpose of identifying the features of language use by TAs in instructional contexts, this paper attempts to first formulate the objectives that an appropriate test of oral proficiency of TAs should pursue and second, explore the ways such objectives can be best translated into task types used for testing the oral ability of TAs. Oral tests constructed in this way, I believe,

can have an excellent washback effect on the teaching that takes place in TA programs prior to the test.

2. The Problem

In recent years, several studies have been performed to identify both the nature of the problem facing TAs and the features of language use they need in instructional contexts. Rounds (1987), for example, analyzed the discourse of effective and non-effective TAs teaching mathematics. She found evidence that when elaborating on mathematics in the classroom, the language of effective TAs included such features as: (a) overtly marking the junctures of problems; (b) explicitly organizing and chunking the material and making transitions between topics; (c) using cohesive devices to relate problems and topics discussed in different sessions. Similar results were obtained by Byrd & Constantinides (1992) who observed native-speaker faculty members--teaching mathematics--who used contextualizing strategies for making connections between classes, between courses, between the discussion and the textbook, between the world of mathematics and the outside world, and so forth. They cite evidence that less successful TAs gave minimal narration of the problems or did not talk at all while solving a problem on the board.

In laboratory contexts, Myers and Douglas (1991) note that the less shared context between students and their TA (compared with that of a teacher-fronted classroom) can be the result of the students' use of non-technical vocabulary and their informal register whose interpretation and comprehension is problematic for the non-native TA. Along with the same lines, in an investigation of TA questions in the chemistry lab, Tanner (1991) reports that the lack of teacher evaluative comments in the lab meant that the language in the lab was more like everyday conversation and concluded that the primary role of the TA within the lab was not to evaluate the students' performance but, consisted of assisting them and giving help.

Office hours constitutes another context within which TAs have to function. In a study of the language used by both professors and TAs with their students in office hours, McChesney (1990) found that the functions of talk included expressing interest and encouragement, recalling what the professor said in class, providing information and advice about tests, and influencing decision making . Questions were used for information, clarification, comprehension, and confirmation. McChesney found that in contrast to professors and TAs, students used indirect forms in asking questions that might not be recognized as questions by non-native TAs.

On the whole, this brief survey of studies dealing with the problem of TAs reveals the type of instructional tasks they need to perform in different academic settings. They refer to three major instructional settings as contexts for language use by TAs: (i) classrooms (one-to-large group interaction), (ii) laboratories (one-to-small group interaction), (iii) office hours (one-to-one or small group interaction).

The nature of all the above contexts for language use is "interactive", emphasizing the importance of teacher questions in organizing classroom talk and the importance of listening comprehension (of both formal and informal registers).

Moreover, the language used for teaching in such interactive contexts is complex in nature, that is , TAs should have the ability to process several language tasks simultaneously (e.g., chunk information, organize it, provide transitional cues, and use appropriate means for best transferring it to the students). As Hoekje and Linnell (1994) state, such complexity can be best captured by communicative models of language such as Canale & Swain (1980) or Bachman (1990).

The above considerations imply that tests designed for assessing non-native TAs' oral proficiency should involve them in the appropriate expression and interpretation of utterances in an interactive setting. The question that arises at this point is whether any one of the present standardized instruments--such as Test of Spoken English (TSE) , Spoken Proficiency English Assessment Kit (SPEAK) , Oral Proficiency Interview (OPI)--which have so far been used in several institutions for assessing the oral proficiency of TAs, really elicit the performance that can be indicative of the TA's ability to employ the interactive complex form of language required by instructional settings.

The next section will focus on this point.

3. Standardized Tests Of Spoken Language

Since its introduction (more than a decade ago), the TSE has proved to be the most commonly used measure of spoken ability. This exam is a tape recorded test in which the examinee's responses are also tape recorded. In an attempt to determine the TSE's validity for testing the language communication skills of non-native TAs, Clark and Swinton (1980) conducted a study using sixty foreign TAs at several large universities in the United States. The study reported a fairly strong correlation not only between TSE and OPI scores (which at the time was assumed to be a valid criterion measure) of the same subjects, but also between the instructors' TSE comprehensibility score and the students' evaluations of the instructors' "ability to handle common situations involving language skills".

Above results, however, contradict Educational Testing Service's (ETS) recommendation that TSE scores should not be considered as the only measures for evaluating TAs and that other available relevant information should also be taken into consideration in addition to TSE scores (ETS, 1990). Moreover, OPI, which consists of a live conversation between the examinee and the interviewer (as opposed to a tape recorder) has recently been strongly criticized by several researchers. For example, Valdman (1988) states:

The total control the OPI interviewers possess is reflected by the parlance of the test methodology . . . In short, the OPI can only inform us of how learners can deal with an artificial imposition rather than enabling us to predict how they would be likely to change authentic linguistic interactions with target language native speakers. (p.125)

Bachman (1988) also questions the validity of OPI on the grounds that it does not distinguish between ability and performance and that its 0-5 rating scale is not theoretically grounded. Furthermore, Hoekje and Linnell (1994) point out that OPI's time-consuming (about 15-25 minutes for each examinee) nature is its greatest limitation.

As for the SPEAK test, it has also been criticized since the type of topics it covers do not adequately assess TA's proficiency as communicators of scientific or technical information (Ponder, 1991). Another area of concern has been the audio-taped format of the test as a means of eliciting language. Students have to talk to a tape recorder about pictures and topics which seem unnatural to them (Gokcora, 1992; Johnson, 1991).

In short, due to the different limitations imposed by these standardized tests, none of the tests can lead to the observation of the examinee's performance in contexts similar to those mentioned at the end of section two. That is why many TA programs nowadays use standardized tests (such as SPEAK) just for initial screening and placement but not for final evaluation. Here is where the test purpose and the actual context of use lead to the requirement of "authenticity" for language tests. Section four elaborates on the concept of "authenticity" as formulated by Bachman(1990).

4. Authenticity: Some Theoretical Considerations

Authenticity has been the topic of major discussions in language testing in the past decade (Language Testing 1985, 2/1) and still continues to be a widely discussed issue in this field. As Bachman (1990) suggests, there are two major approaches to authenticity. The "real life" approach requires tests to mirror the reality of non-test language performance. It is mainly concerned with the appearance of a test and the way it affects the test performance (face validity); and the extent to which test performance predicts future non-test performance (predictive validity).

The "real life" approach--which underlies the Interagency Language Roundtable (ILR) oral interview (OPI),--one of the most widely used approaches to testing speaking--has been criticized on the grounds that the replication of

non-test performance in language tests is not possible and no test situation can exactly resemble its real life counterpart (Spolsky, 1985). For example, there are major differences between a situation in which a lecturer presents information to undergraduates in a regular session during a course and when he presents information to a panel for certification purposes (Hoekje and Linnell, 1994).

The second approach to defining authenticity, called the "interactional/ability" approach, focuses on the distinguishing characteristics of communicative language use; the interaction between the language user, the context, and the discourse. Tests based on this view involve the test taker in the appropriate expression and interpretation of illocutionary acts, and his performance in the test is interpreted as an indication of the extent to which he possesses various communicative language abilities. As opposed to the real life approach which does not distinguish between the behavioral manifestation of an ability and the trait itself, here there is a clear distinction between the abilities to be measured, on the one hand, and the performance we observe and the context in which observations take place, on the other. In the interactive/ability (IA) approach, the role of context and the manner (i.e. test method facets) through which we elicit a sample of performance are, therefore, recognized. This would result in a shift of emphasis from "attempting to sample actual instances of non-test language use to that of determining what combination of test method facets is more likely to promote an appropriate interaction of a particular group of test takers with the testing context" (Bachman 1990, p.317). According to this view, authentic or communicative tests or test tasks should reflect our knowledge of language abilities and language use.

On the whole, Bachman's model of language testing consists of two interrelated theoretical frameworks that affect performance on language tests: "communicative language ability" and "test method facets". In his formulation, communicative language ability consists of language competence, strategic competence, and psychophysiological mechanisms. Test method facets, on the other hand, refers to five main features of a test: (1) the testing environment (e.g., familiarity of place and equipment), (2) the test rubric (e.g., organization, time allocation), (3) the nature of input the test taker receives (format and the nature of the language), (4) the nature of the expected response to that input (format, the nature of language, and restrictions on response), (5) the relationship between input and response (e.g., reciprocal or non-reciprocal).

5. Ia Authentic Tests And Tas Evaluation

Within the "communicative language ability" component of Bachman's framework for language testing, differences in language ability can be attributed to differences in language knowledge and/or in the adoption of communication strategies. Such a notion can also account for the communicative competence necessary for efficient teaching assistants. Their ability to handle complex functions of language in such interactive contexts as instructional settings can thus be tested by making use of a criterion-referenced performance test the activities of which best involve TAs in communicative language use. Specifications for the development of such a test can be characterized on the basis of Bachman's theoretical framework in the following way.

First, with respect to the "communicative language ability", the test should be designed so that it captures:

1. Language competence: the oral test should tap TA's knowledge of vocabulary, grammar and morphology to the extent that these elements interfere with the communication of information. Moreover, TA's performance in the test should involve the employment of elements of conversational language--such as the use of a variety of cohesive devices for organizing the oral discourse--as well as the employment of conversational conventions such as attention getting, topic development, communication maintenance (Hatch, 1978). Besides the above elements of organizational competence, the examinee's pragmatic competence including knowledge of language functions and sociolinguistic rules of appropriateness should be also considered as an important element to be measured by the test.
2. Strategic competence: the degree to which the test taker adopts communication strategies to overcome his/her linguistic gaps in interactional situations is the most important point that the

test should measure in the performance of TAs. Successful communicators try to achieve their communicative goal by resorting to different types of communication strategies (such as paraphrase, circumlocution, repetition, overgeneralization, . . . etc) rather than avoiding communication altogether when there is a gap in their linguistic knowledge.

3. Psychophysiological mechanisms: the implication of this factor for a TA oral assessment test is that the test should involve the test taker not only in productive but also receptive language use. For instance, the TA should use his/her auditory skill to gain access to the information on the tester's questions/instructions, and his/her articulatory skill to pronounce the word correctly.

The second dimension of the model, "test method facets", affects the examinee's performance on language tests and can be adopted for our purpose as follows:

1. Testing environment: Tests should be conducted in a familiar place with familiar equipment (the ones usually used by TAs in their teaching activities such as blackboard, opaque projector, etc). Because of the interactive nature of the test, no tape recorder is used, rather test takers should address live audience.
2. Test rubric: Test takers are instructed to use English during the testing session, no native words are permitted. It is very important that the test taker understands the task that is required of him/her (through aural channel). Timing is relatively fixed regardless of whether the teaching task performed by the test taker is finished or not. Within the specified guidelines, it is the test taker who determines the content area and its organization of parts.
3. Input: The input given to the test taker is in the form of "live" language from several different audience members by making use of the target language (English in this case) through aural channel.
4. Response: The type of the response should be "constructed", i.e., consists of the language sample in response to the input material. Regarding the nature of the language of response, the test should involve the TA in using highly specialized vocabulary associated with technical registers as well as references to figures of speech which often characterize effective language use. As for the topic of the talk, the interactive relationship between input and the response requires that the topic of the discourse be determined jointly by the tester and the test taker so that a topic which is best suited to the candidate's interests and knowledge is chosen. Examinee's talk around this topic takes place in a limited time during which his/her employment of different language functions as well as the listening comprehension are explicitly evaluated.
5. The relationship between input and response: This relationship is reciprocal necessitating the presence of feedback on the part of the examiners. Thus both the test taker and the audience have control over the conversation and turn-taking.

6. Final Remarks

The above detailed specifications reflect how the model proposed by Bachman can be adapted to a criterion-referenced performance test with the specific purpose of assessing the oral proficiency of teaching assistants.

An important consequence of devising a test on the basis of the above procedure, I believe, is the positive washback effect that it can have on instruction by reflecting the skills and abilities that should be focused upon in

any TA preparation course. It is then quite clear that the contents of a language course for TAs with a focus on developing the oral ability should be different from a course in conversation with an oral interview as its final test. The elements of communicative performance that exist in a performance test devised on the basis of the above specifications provide feedback for teaching and learning by bringing particular needs of teaching assistants to the attention of those responsible for material development and course design.

References

- Bachman, L. 1988. "Problems in examining the validity of the ACTFL oral proficiency interview". Studies in Second Language Acquisition. Vol. 10/2: 49-164.
- Bachman, L. 1990. Fundamental Considerations in Language Testing. Oxford: Oxford University Press.
- Bailey, K. 1984. "A typology of teaching assistants". In K.M. Bailey, F. Pialorsi, & Zukowski/Faust (eds.). Foreign Teaching Assistants in U.S. Universities. Washington D.C: NAFSA.
- Brown, G., & G. Yule. 1983. Teaching the Spoken Language. Cambridge: Cambridge University Press.
- Brown, H.D. 1993. Principles of Language Learning and Teaching. Englewood Cliffs, NJ: Prentice-Hall.
- Byrd, P., & Constantinides, J. 1992. "The language of teaching mathematics: implications for training ITAs". TESOL Quarterly, 26/1: 163-167.
- Canale, M., & M. Swain. 1980. "Theoretical bases for communicative approaches to second language teaching and testing". Applied Linguistics, 1: 1-47.
- Clark, J.L.D., & S.S. Swinton. 1980. "The test of spoken English as a measure of communicative ability in English-medium instructional settings". TOEFL Research Report 7. Princeton, NJ: Educational Testing Service. Educational Testing Service. 1990. Test of Spoken English: Manual for Score Users. Princeton, NJ: Author.
- Gokcora, D. 1992. The SPEAK test: "international teaching assistants' and instructors' affective reactions". Paper presented at the 26th Annual TESOL Convention, San Francisco.
- Hatch, E. 1978. "Discourse analysis and second language acquisition". In E. Hatch ed.) . Second Language Acquisition: A Book of Readings. Rowley, Mass.: Newbury House: 401-35.
- Heaton, G.B. 1990. Writing English Language Tests. London: Longman.
- Hoekje, B., & K. Linnell. 1994. "Authenticity in language testing: evaluating spoken language tests for international teaching assistants". TESOL Quarterly, 28(1), 103-126.
- Hughes, A. 1991. Testing for Language Teachers. Cambridge: Cambridge University Press.
- Johnson, K. 1991. "Modifying the SPEAK test for international teaching assistants". TESOL Matters : 8.

- Lynch, B. K., & F. Davidson. 1994. "Criterion-referenced language test development: linking curricula, teachers, and tests". TESOL Quarterly, 28 (4), 727-743.
- McChesney, B. 1990. "University office hours: what professors and teaching assistants say to students". In G. Barnes, M. Berns, & C. Madden (chairs). Training of International Teaching Assistants. Preconference symposium presented at the 24th Annual TESOL Convention, San Francisco.
- Myers, C., & D. Douglas. 1991. "The ITA lab assistant: strategies for success". Paper presented at the Annual NAFSA Convention , Boston.
- Ponder, R. 1991. "The TSE: a language teacher's critique". Paper presented at the 25th Annual TESOL Convention, New York.
- Rounds, P. 1987. " Characterizing successful classroom discourse for NNS teaching assistants". TESOL Quarterly, 21 (4), 643-671.
- Spolsky, B. 1985. "The limits of authenticity in language testing". Language Testing , 2 (1) , 31-40.
- Tanner, M. 1991. "NNSTA-student interaction: an analysis of TA's questions and students' responses in a laboratory setting". Unpublished Doctoral Dissertation, University of Pennsylvania, Philadelphia.
- Underhill, N. 1989. Testing Spoken Language. Cambridge: Cambridge University Press.
- Valdman, A. (ed.). 1988. "The assessment of foreign language oral proficiency". Studies in Second Language Acquisition, 10 (2).

