

Considerations in Classroom Testing

Rebecca Tanaka and Jim Middleton

Asia University and Japan Women's University

Classroom testing and student evaluation are important parts of any language program. Not only do test results allow the teacher to accurately determine grades for the students, but they also provide feedback and guidance to the teacher regarding lesson planning and methodology. Additionally, grades and types of tests send signals to the students about how to study and what is required of them in class. Although testing is often the last part of the course-planning process, it is one which deserves careful consideration.

In this paper we will review three key areas to be considered in test design and methodology: (a) proficiency and achievement, (b) test format, and (c) test validity. We will recommend a combination proficiency/achievement test style and will present a schemata which can be used by teachers to evaluate their test designs.

Proficiency and Achievement

By definition, there is a difference between proficiency tests and achievement tests. Proficiency tests are designed to measure general progress in the acquisition and use of language skills and are not meant to assess learning from any individual course syllabus (Dandonoli, 1987). Achievement tests by contrast are more diagnostic and test specific materials that have been introduced to students during a particular course (Clark, 1972).

In examining the English Language Education Research Institute (ELERI) program, the focus on proficiency has had and continues to have a strong effect on student placement,

curriculum design, materials development and teaching methodology. Students are placed into ability level groupings on the basis of their scores on listening and reading proficiency tests. Student-learning goals are written in terms of what students should be able to do in the target language. Classroom activities frequently focus on using English to communicate ideas and information. Similarly, classroom tests must reflect this focus, or the tests may create a weakness in the overall language program.

If tests in a program focus on discrete-point knowledge exclusively, students receive the signal that striving for proficiency in the target language is not important because it is not tested (Bacon and Finneman, 1990). Discrete-point knowledge, in this case, refers to single word matching to indicate knowledge of vocabulary or noncontextualized grammar-point questions.

Because of the purpose and range of materials being tested with achievement tests, there is a great tendency for teachers to use a discrete-point style. This is because from a teacher and a student point of view, there are many advantages to using a discrete-point achievement test format. It is more straightforward, accessible and objective. Also, teachers and students are familiar with this type of test and feel comfortable with it. Furthermore, discrete-point achievement tests are less time-consuming and easier for teachers to design and to grade than discourse-based and functional tests; discrete-point tests are also easier for students to study for.

However, by utilizing a discourse-based and functional test format, classroom tests can be designed to reflect proficiency

goals and measure achievement by presenting language in context. In this type of test, students are required to use the target language beyond sentence level to carry out realistic tasks, student proficiency is encouraged, and the scope of material is limited to that which was introduced in the segment of the course being tested. Also, because a combination proficiency/achievement test steers away from single-sentence test questions, teachers can more accurately assess students' communicative competence (Wesche, 1981). The following passage sums up this point:

Language testing which does not take into account propositional and illocutionary development beyond the sentence level, as well as the interaction between language behavior and real-world phenomena, is at best getting at only a part of communicative competence. Small wonder that we often find that a student's success at second-language classroom exercises and tests appears to bear little relationship to his or her ability to use the language effectively in a real-world situation (Wesche, 1981, pp. 552-53).

Therefore, a hybrid proficiency and achievement test can produce the most desirable results for a program similar to ELERI's, a communicative, content and function-based program which has general student learning goals and objectives.

Test Format

In designing a proficiency/achievement test, the teacher's task is to create an examination requiring students to show how well they know specific features of English and to demonstrate their abilities to use the language in natural discourse. Grammar, context, structure, and situation must be combined (Slager, 1978). A test which uses both open-ended (divergent) responses with specific (convergent) items is desirable. For example, some sections of the test can focus on discrete points

of grammar, vocabulary, discourse or pragmatic features, while other sections allow students to respond more freely, using whatever language they know to complete the task (Hadley, 1993).

In order to more fully explain the proficiency/achievement test style, refer to the following sample discrete-point achievement test and the continuums for assessing characteristics of test items. For additional discussion of this concept and examples of proficiency/achievement test formats, refer to Hadley, 1993.

A. Write an adverb of frequency to complete the following sentences. Use a different adverb of frequency for each sentence.

1. I _____ get up at 7:00 o'clock.
2. I _____ play soccer.
3. When I was in high school, I _____ did my homework.

B. Write the correct vocabulary word in the sentences below.

1. My parents pay my _____ to university.
2. In American universities, the first _____ begins in September each year.
3. Many American students live in _____ while attending university.

C. Fill in the blanks with the correct form of the verb.

1. It's necessary that they _____ see me. (to come)
2. It is possible that I _____ at home this evening. (to be)
3. It is possible that we _____ some money. (to have)

Please notice that all test questions are single sentence items which require insertion of one piece of information. There is no connecting context as there would be in natural discourse.

If a teacher chooses to use these kinds of test questions, global comprehension-type questions should also be included (Hadley, 1983). By combining the two types of questions, teachers can more accurately gauge achievement and proficiency.

When designing and reviewing the test prior to administering it to students, the following continuums are useful to teachers in assessing the balance of questions. The left side of the first continuum represents questions like those in the sample discrete-point test above. They measure language recognition but give no indication that students can use the language for communication. Therefore, teachers should strive to have few, if any, questions that fall on the left side of the continuum and have more global, language production type of questions represented on the right side of the continuum. The global or open-ended type of questions allow teachers to measure how well and how much students can use the language for communicating thoughts and ideas.

Measures language
recognition

Measures
communicative
ability

Discrete-point
questions (only
one correct answer)

Global comprehension
questions (many
possible answers)

The continuum below is used to assess the balance of single-sentence or phrase test items and sequential, natural discourse items. Items on the left would include drill-like, *textbook* language that does not resemble genuine language use while those items falling along the right side could be excerpts from natural discourse or should approximate authentic communicative exchanges (Hadley, 1993). Since natural language occurs in context, testing students with noncontextual questions, which fall on the left side of the continuum, creates a false reality. Testing which contains natural styles of communication and language give a more accurate evaluation of students' language abilities and knowledge.

Less Desirable

More Desirable

Series of unrelated single sentences or phrases

Sequential, natural discourse material

To use this test design and review process, if past-tense narration were being tested, for example, discrete features could include past-tense verb forms, adverbial connectors, time expressions, etc. The teacher could design several test questions to determine how well students could use the function of narration, ranging from formats using specific linguistic or lexical features to those with free responses. A cloze passage could be used in which specific parts of speech, such as verbs in the past tense, had been deleted and students falls in the gaps using the verb clues. This type of format would fall on the left side of the horizontal axis and demonstrates knowledge of a discrete-point nature. Items in which the students have to write a short paragraph telling what they did during New Year's break and using the appropriate past tense verb forms falls along the right side of the horizontal axis. Such a format demonstrates more global knowledge of the language and proficiency.

The context for the test items would be selected from the content themes addressed in the class plus functional and communicative items. The test writer must be particularly explicit about what language skills are to be tested. In addition, the teacher must design the test so that the communicative language ability being tested is as similar as possible to the skills and materials taught and practiced in class. Many of the test items should ideally be on the right side of the natural language continuum for the test to best

represent real-world language with at least a portion of the questions on the right side of the discrete-point/global knowledge continuum to gauge student proficiency. When a test is written like the previous sample discrete-point achievement test above, the test items all fall on the left side of the discrete-point/global knowledge continuum. This type of test would not represent ability to use English and therefore would not be an indication of student ability to synthesize the pieces of language being tested.

Below is an example of a listening comprehension test in which students listen for specific information. The context is a radio sports broadcast. The tape script replicates natural speech patterns and information is often repeated more than once. This test would fall on the far left side of the discrete-point/global communication continuum and on the far right side of the single sentence/natural discourse continuum.

Context: A radio sports broadcast

Function: Listening for specific information

Student task: Students listen to a simulated sports broadcast in which the starting line-up of a student basketball game is being announced. As they listen, they place the number of the player next to his name.

_____ Bill Davis
_____ John Murphy
_____ Tony Valencia

_____ Dan Mendoza
_____ Ryan Carter

Listening Passage: Now let's bring out the visiting men's basketball team from Central Washington University. First we have number 5, Bill Davis. Bill Davis, number 5, is from Spokane, Washington. From Cloverdale, Oregon, comes number 10, Ryan Carter, number 10, Ryan Carter from Cloverdale, Oregon. Also from Oregon comes Tony Valencia, number 13. Number 13, Tony Valencia from Oregon State. And from Idaho comes number 20, John Murphy, Mr. Murphy, number 20. Finally, from Seattle, Washington, comes the fantastic Dan Mendoza, number 35. Yes, fans, the famous Dan Mendoza, number 35.

On other tests, the far left, convergent, discrete-point test item *listen for specific grammatical or lexical features* might require students to listen for cues to tense, gender, or number, or, if the embedded cues are lexical, might require students to listen for numbers, colors, body parts, etc. For upper-intermediate and advanced students, the teacher would play a tape or read the short passage once. For lower and lower-intermediate ability levels, the passage would be read again more slowly, pausing so students could write their answers. Natural language is often repetitive and offers multiple cues to meaning. Although students would listen for discrete features of the discourse, they would hear the passage in context, which more closely resembles natural language. In this style of test, students could use nonmorphological data such as time words or situation to help them determine the correct answer (Hadley, 1993).

The following continuum lists various types of listening activities. Test items that require students to report the gist of a listening passage or to classify it globally in some way would fall on the far right side. Students could listen to a passage and either choose from multiple-choice options that would best paraphrase the main idea of the listening or give it a title using either multiple choice options or free creation, depending on the ability level of the students (Hadley, 1993).

Convergent
Discrete-point
Achievement

Divergent
Global
Proficiency

*	*	*	*	*
Listen for specific grammatical or lexical features	Listen for specific semantic information	Listen and follow map, diagram; choose picture	Write summary of message content in native language	Choose appropriate title, summary statement

One objection to including the global or free creation, proficiency-based type of test items such as those described above right is that they are more difficult to grade. One suggestion is that this type of test format may be scored in either a global or a discrete-point fashion.

points could be awarded for the comprehensibility and quality of the communication in a general sense, or they could be awarded for particular features of the message (such as the appropriate use of the tense and the correct forms of the verbs.) A scoring scheme that awards credit for both the general comprehensibility of the answer and specific discrete features is another viable option. (Hadley, p. 420)

In summary, a combination of test items ranging from discrete-point to more divergent and global comprehension offers the optimum format for proficiency and achievement testing. This format signals to students that they cannot merely memorize vocabulary or grammatical forms and do well on the test.

Test Validity

The third area for teachers to consider when developing assessment tools is test validity. Test validity is a number of basic subsets consisting of content validity, construct validity, and face validity (Hughes, 1989). For a test to have high content validity, the language skills, structures, or information taught in the class should be represented on the

test instrument. Then the teacher should include all learning areas taught in the unit being tested and specify to the student exactly what is being tested. When learning areas are unclear, underrepresented, or not represented on tests, not only is there low content validity, but also the *negative feedback effect* can develop. In other words, those areas which are not tested tend to be ignored in future teaching and learning interactions between teacher and student (Hughes, 1989).

The second subset of test validity is construct validity which refers to whether the test instrument actually measures what it is designed to measure. Construct validity is often overlooked unless the test designer or proofreader reviews test questions from a highly critical point of view. In recent years an accepted premise in EFL reading methodology is that the meaning of an individual vocabulary word, expression or phrase can be guessed by the student from the context in which the unknown word is encountered. If the teacher attempts to measure students' abilities to guess meaning from context, the test must clearly demonstrate that the specific ability of guessing meaning from context was being measured (Henning, 1987). There are numerous variables involved in this example such as some students knowing the word and other students randomly guessing rather than guessing from context. These variables make it complex to test an ability like guessing meaning from context. By asking a colleague to review a test prior to administering it to the students, construct validity questions can more easily be avoided.

The third type of test validity is face validity. If a test looks like it measures what it is supposed to measure, it

is said to have face validity (Hughes, 1989). For example, a pronunciation test in which students are not required to speak would lack face validity even if it had content and construct validity. This element of validity is important from a psychological perspective; students who perceive that the test lacks face validity may not perform according to their true ability (Hughes, 1989).

In summary, for a test to be considered valid it must measure what it intends to measure. It must also focus on the particular skill or skills that have been taught in the lessons and practiced in the classroom. Furthermore, only data specified in the test instructions should be evaluated.

Additional Resources for Student Assessment

Creating tests which accurately measure students' language abilities and proficiency is not a simple task. To aid in assessing student progress and assigning grades in a communicative style course, particularly with 25 plus students per class, the range of assessment procedures should be broadened. In addition to valid proficiency/achievement style tests, teachers may find the following types of graded work useful. Collect samples of students' work in portfolios, take samples of oral performance during classroom activities, video or audio tape role-plays or presentations, and conduct oral interviews. All of these evaluation techniques measure student proficiency in the target language. By adding these types of graded activities, the teacher can increase the content, construct, and face validity factors of the overall assessment process as well as give students a variety of situations in

which to demonstrate their abilities to use the English language.

References

Bacon, S. M. & Finneman, M. D. (1990). A study of the attitudes, motives, and strategies of university foreign-language students and their disposition to authentic oral and written input. *The Modern Language Journal* 74, 459-73.

Clark, J. L. D. (1972). *Foreign language testing: theory and practice*. Philadelphia: Center for Curriculum Development.

Dandonoli, P. (1987). ACTFL's Current Research in Proficiency Testing. In H. Byrnes and M. Canale (Eds.), *Defining and developing proficiency: guidelines, implementations and concepts* (pp. 75-96). Lincolnwood, IL: National Textbook Company.

Hadley, A. O. (1993). *Teaching language in context*. Boston: Heinle & Heinle Publishers.

Henning, G. (1987). *A guide to language testing, development, evaluation and research*. Rowley, MA: Newbury House.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Omaggio, A. C. (1983). *Proficiency-oriented classroom testing*. Washington, DC: Center for Applied Linguistics.

Slager, W. R. (1978). Creating contexts for language practice. In E. Joiner and P. Westphal, (Eds.), *Developing communication skills*. Rowley, MA: Newbury House.

Wesche, M. B. (1981). Communicative testing in a second language. *Canadian Modern Language Review* 37, 552-553.