Diagnostic Language Testing for Taiwanese University Students: The Online English Assessment System (OEAS) Project
Muchun Yin and James Sims
Tunghai University
muchun.yin@thu.edu.tw

Diagnostic language testing aims to help test-takers better understand their own linguistic strengths and weaknesses so as to aid language learning. While the goal is admirable, there are relatively few truly diagnostic language tests currently in existence, the main and perhaps best example being the DIALANG Project (Alderson, 2005). This paper presents a synopsis of a diagnostic test construction project currently in its early stages at a Taiwan university, the Online English Assessment System (OEAS). The authors first describe the background to the OEAS, and then outline its main design features. In addition, one important characteristic of the OEAS is its “local” nature: it is being produced by some of the university’s EFL teachers for use by the university’s students as a self-assessment tool, and thus it is also expected to be linked to the university’s language learning resources. The authors also highlight key areas for the project’s validation work.

Key words: diagnostic language test, test validation, test construction, self-assessment

Introduction

In 2005, the Taiwan Ministry of Education (MOE) began a four-year initiative to fund projects at universities around the island for the purpose of promoting teaching excellence in higher education. Tunghai University proposed and was awarded money for six such projects. Under one of these projects was the Online English Assessment System (OEAS) sub-project, whose goal is to construct and validate a diagnostic language test that will assist the university’s students in their language learning. In this paper, we briefly introduce diagnostic language testing, highlighting its importance and primary characteristics. This is followed by a short history of the OEAS project, focusing particularly on decisions about test purpose and use, and a sketch of the test’s main design features. After that, two aspects of the test’s “local” nature are discussed: the teachers who are involved with test development, and the university resources that are to be connected to the OEAS. The paper concludes with a discussion of three areas of emphasis in the test’s validation program.

Diagnostic Language Testing

The primary characteristic of a diagnostic (foreign) language test is that it identifies a test-taker’s specific linguistic strengths and weaknesses (Alderson, 2005; Bachman & Palmer, 1996).

This can be contrasted with proficiency and placement tests, which often label a test-taker with a general level of language ability (e.g., low, intermediate, or advanced) and/or match a test-taker with an appropriate language course (perhaps based on a general level, or on a level in a general skill area like speaking), but do not provide meaningful feedback to the test-taker or test-giver about particular aspects of language that need further improvement (for example, that the test-taker is weak in particular grammar points like past tenses or adjective clauses, or in particular listening or reading sub-skills like listening for main idea, reading for specific information, or drawing inferences). It is this provision of meaningful feedback to test-takers about their language abilities that is the most important and useful aspect of diagnostic testing, and also perhaps its most challenging aspect in terms of test construction.

Diagnostic language tests can also be contrasted with achievement tests, even though both types deal with specific linguistic strengths and weaknesses. An achievement test usually occurs during or at the end of a course, and often deals only with aspects of language ability that are covered in the course. A diagnostic test, on the other hand, may serve as a useful complement to a course but can occur independently of a course and is not limited by
that course’s content in terms of coverage. One distinct advantage of this is that diagnostic tests are thus well-suited for aiding the self-assessment of language ability, helping learners who perhaps are not currently enrolled in a language course to focus their self-study upon particular areas of weakness.

In what so far seems to be the only in-depth book on diagnostic language testing, Alderson (2005) points out that while the idea of diagnosing foreign language ability has often been mentioned in the testing literature, there are very few truly diagnostic tests in existence. The primary, if not only, instance of such a test is DIALANG, a diagnostic test sponsored by the European Commission (which Alderson and his colleagues have been working on since the mid to late 1990s, and which forms the basis for his aforementioned book).

This brief overview is meant to highlight both the importance and relative newness of diagnostic language testing. In the next sections, we give a synopsis of one diagnostic test construction project currently underway at a Taiwan university.

Background of the OEAS Project

In early 2005, the MOE invited proposals for so-called “Teaching Excellence” projects from qualified universities around the island. In the Foreign Languages and Literature Department (FLLD) at Tunghai University, the chair and several other faculty discussed, developed, and proposed several projects that were later approved for funding. One of these was an online English language test, later dubbed the Online English Assessment System (OEAS). When the test was proposed, the main thought was that the test’s purpose was to give students some bearing on their language ability level in comparison to tests like the General English Proficiency Test (GEPT), especially as increasing numbers of higher education institutions and their departments, not to mention a wide variety of employers, are requiring certified levels of English language ability. Later, as the members of the project team—nearly all of whom are FLLD faculty—held further discussions about test purpose, direction, and content, two things became clear. First, we were not interested in reinventing the testing wheel; we did not want to produce another summative high-stakes proficiency test like the GEPT, TOEFL, or IELTS. We also were aware of the work on NETPAW, the online proficiency test being developed at the Open University of Kaohsiung, and did not want to overlap with that. Second, we were all teachers, our department emphasized teaching, and our university had a long tradition in providing a liberal arts education, so we wanted to develop a test that had pedagogical value, not merely label students with a score. In addition, one of the team members during his recent doctoral studies had become aware of the growing research into “assessment for learning” or formative assessment (Black & Wiliam, 1998) generally and into diagnostic language testing specifically. The latter seemed very well-suited for the pedagogical purpose that the team had in mind, and the dearth of diagnostic tests—besides DIALANG, which while an admirable exemplar is nevertheless designed for the European context rather than the Asian/Taiwanese context—seemed to signal that the OEAS had found its niche: to provide a diagnostic online English language test for Tunghai students.

OEAS Design Features

Like DIALANG, which provided our team with an important model for diagnostic test design, the OEAS is actually planned as a battery of tests. There are to be two main sections: a general test and a group of skills tests. These are described presently (further information can be found at the OEAS website: http://www2.thu.edu.tw/~langexam).

The General Test
Students will first take a general proficiency test that features 60 multiple-choice questions—20 each of grammar, listening, and reading—and is expected to measure students’ proficiency in general English. This first test, which is actually a retired freshman English placement exam developed by our department, will have a general “macro-diagnostic” (Alderson’s [2005] term) function; that is, it will give students an idea of their general language level. Two features are expected to make this test fulfill its diagnostic (as opposed to placement or proficiency measurement) purpose. One is that it will give students an idea of how they would fare on other more high-stakes summative tests like the GEPT, TOEFL, IELTS, and TOEIC. These rough estimates will be based on research comparing the OEAS general test with these other tests. For example, a correlation study was conducted in January 2006 comparing the high intermediate GEPT with the OEAS general test (discussed below). Another feature is that it will provide test-takers with feedback on their incorrect answers after they have finished the test. This is important to highlight the pedagogical use of testing, and to encourage test-takers by suggesting that they can always learn more to improve their language ability.

The Skills Tests

After a student has taken the general test, he/she will have the option to take one or more multiple-choice tests in specific aspects of language: grammar, listening, and reading. These tests will give the student specific “micro-diagnostic” (again, Alderson’s term) information about the student’s strengths and weaknesses in that linguistic area. For example, the grammar test will provide information about which grammar structures (e.g., past tenses, coordinated clauses, articles) the test-taker is weak or strong in, and the listening and reading tests will do the same for listening and reading sub-skills (e.g., listening for appropriate response, reading for main idea, guessing meaning of words in context).

Each structure or sub-skill requires several questions in order to give some idea of the test-taker’s grasp of it, so the number of questions required per test are as follows:

- The grammar test: 15 structures x 4 questions per structure = 60 questions
- The listening test: 8 testlets (Buck, 2001), with each testlet containing a listening passage (4 involving an individual speaker, 4 involving two speakers) and a question for each of the 3 listening sub-skills = 24 questions
- The reading test: 4 testlets, with each testlet containing a general expository reading passage and 8 questions covering the 5 reading sub-skills = 32 questions

At this point, the skills tests are to be aimed at an intermediate ability level, defined here as students who are able to score 60% or higher on the general test. It is hoped that as resources permit, another set of more difficult skills tests will be created.

After taking each exam, the student will receive several kinds of feedback (both in English and in Chinese):
1) a summary report will be given that lists percentages correct for each structure or sub-skill; this is meant to let the test-taker know which aspects they are strong and weak in.
2) the student can click on each structure or sub-skill label to receive advice on specific sections of current textbooks that can give further information and practice.
3) as with the general test, students will also receive specific item feedback (e.g., explanation of the relevant grammar rule) on questions answered incorrectly.

These forms of feedback are key elements if the OEAS is to fulfill its diagnostic purpose, and therefore much of the project’s focus will be on crafting feedback that students can understand and use.

The Local Nature of the OEAS
One distinctive characteristic of the OEAS is that it is very “local” in two important ways: first, the members of the test construction team are all teachers in the university; and second, the test is to be connected to the university’s own resources, particularly its English course offerings and its self-access language learning center. Issues arising from these are discussed below.

Teachers as Test Creators

The project team currently consists of 9 teachers in the FLLD. There are some disadvantages of this. One is that, as university faculty, they simply do not have much time to work on the project, as they need to fit project work into their busy schedule of teaching and administrative responsibilities. Another disadvantage is that, as with teachers generally (Stiggins, 2001), there is a lack of knowledge relating to testing theory and research, and so there is unfamiliarity with aspects of diagnostic and online testing.

However, there are also distinct advantages of having teachers working on the project. First, they have a great deal of practical experience with test construction and item writing. The midterm and final exams for many of their courses are coordinated, so they are used to working in teams to produce quality tests. The second advantage—vital for diagnostic test construction—is that they are familiar with Taiwanese students’ linguistic strengths and weaknesses; this knowledge is then expressed in test items and feedback that are aimed specifically for the university’s students. For example, some distractors may be ones that the teachers know many students will be attracted to, and the feedback can explain why those distractors are incorrect.

There have been some interesting “side effects” resulting from this relationship between the OEAS and the teachers. One is that working on the project has allowed teachers to make explicit some of their implicit knowledge or assumptions (Borg, 2003) about item writing and testing. For example, one teacher had a rule of thumb for grammar questions—“don’t use ungrammatical constructions (e.g., ‘he sleeped at night’) as distractors”—that he was not sure should be used in this test because of its pedagogical nature (the thought being that many students make similar ungrammatical mistakes and thus could benefit from having those mistakes highlighted), so he had to articulate it and a decision had to be made as to whether such a rule was to apply to OEAS item construction. Another intriguing effect has been that teachers are revisiting aspects of the tests they use for their classes in light of project discussions. For instance, one section of the midterm and final exam for the school’s first-year English course involves listening to a story and then answering some questions. During a project meeting, a question was raised as to whether such an arrangement actually tests memory more than listening ability, because the OEAS listening test takes a different approach. One can therefore see that, by providing opportunities for reflection and discussion about exams, the OEAS project can serve as an avenue for teachers’ professional development in the area of language testing.

Connecting to University Language Learning Resources

Another important sense in which the OEAS is “local” is that it is to be linked to the university’s language learning resources. One such resource is the self-access language learning center, which is another Teaching Excellence sub-project currently underway. The two actually complement each other; students can take the OEAS to get a bearing on their English strengths and weaknesses, and then access suitable resources in the center for their self-study. The key connection will be in the test feedback to students. Instead of vague or general advice like “watch more movies in English” or “listen more to the English radio station,” students will be pointed to specific resources (e.g., books, magazines, multimedia packages, and websites) that are appropriate for their level and weaknesses.
Another language learning “resource” is the university’s English teachers (and the courses that they offer). For the OEAS to be more than an interesting but irrelevant technological relic/”white elephant,” it is essential to find ways to make it useful to teachers and to integrate it into English courses at the university. This may be more difficult than actually creating the test. Last year, we conducted a survey of full-time FLLD faculty who had taught on the university’s first-year English course; we asked if and how they used exam results from the course placement, midterm, and final exam. Only a few teachers responded, which indicated to us at least some apathy and indifference towards this topic, and the ones who did respond mostly said they rarely used those exam results. At least the FLLD teachers who responded to our survey were positive when asked about the idea of a diagnostic test of strengths and weaknesses in language ability. In a sense, any new test is an educational innovation (Wall, 1996); from this perspective, the OEAS project’s task is not only to construct a diagnostic test but also to try and manage its uptake as an innovation (this also relates to validation research into the consequences of test use, discussed below).

Foci for Validation

In our validation work, we take as our starting point an evidence and argument-based approach to validating the interpretations and actions based on test scores (Kane, 1992; Messick, 1989). Time and space do not permit an extended discussion of our validation strategy; instead, we only wish to highlight below a few areas we see as foci for our validation research.

One area is of course the validity of the constructs tested in the OEAS. Our program here includes:

- correlating our test battery with other tests of similar constructs;
- having expert judges check content validity of the tests (as support for construct validity);
- checking reliability of our tests (also as support for construct validity).

We have already done some work in this area, especially with our general test (see also Sims, forthcoming):

- A comparison of the test with the high-intermediate GEPT using Spearman’s rank correlation analysis produced the following results.
  
  - OEAS General Test and GEPT total scores 0.7653, p<0.0001
  - OEAS General Test and GEPT listening scores 0.7046, p<0.0001
  - OEAS General Test and GEPT reading scores 0.6079, p<0.0001

- Teachers not involved with the test’s construction compared test specifications and test content and agreed that there was a good match between them.

- A split-half reliability coefficient of 0.88 was calculated, thus showing reliability of the general test to be high.

Another area for our research is the various kinds of feedback that are to be given to the test-taker. At this point, we are making an assumption that the types and amounts of information being given to the student after he/she completes the OEAS battery will be usable and beneficial to the student. For example, we assume that the explanations of items that a test-taker answered incorrectly will help that student improve in the tested structures or sub-skills. We also assume that providing estimated scores on tests like the GEPT, albeit rough and crude, will still help students plan their preparation for those tests. However, these assumptions need to be supported (or not) empirically. Here, we think qualitative methods like think-aloud protocols (Ericsson & Simon, 1993; Gass & Mackey, 2000) and semi-structured interviews can provide us with that empirical data.

A third area of focus for our validation work is in the consequences of OEAS test use. Here, one essential question is whether and how students take further action in light of their
OEAS test results. Another important question is whether and how the OEAS can have positive washback effects upon the university’s language teaching. These are daunting questions for our project, because answers to these questions not only will serve as evaluations of the OEAS, but also may show that a much greater and laborious task is required; our research in this third area may indicate that getting people to use the OEAS (which is low-stakes and primarily for self-assessment) will require changing many students’ attitudes towards language testing and language learning—e.g., from passive to active, from externally motivated to internally motivated—and thus will likely involve something of a cultural change, not only test construction.

Conclusion

In this paper, we have outlined the background, features, and validation emphases of a diagnostic language test currently under construction at a Taiwanese university. As we hope this paper has shown, the challenges are many—we have a lot of work ahead of us—but the potential benefits are obvious: students will gain a more detailed portrait of their linguistic strengths and weaknesses, and will be able to receive specific advice and suggestions to guide their own language learning. If this is realized, then perhaps the OEAS can serve as an example of a truly pedagogical language test—a test that encourages and assists language learning.
References


