

Developing Useful Tests for Japanese Students

Shingo Morito

Introduction

With more emphasis of English education placed on nurturing the ability to communicate in English rather than knowing about the language, many teachers are struggling to find the best way to help the students acquire that ability and as teachers our responsibility is to set a clear goal for the students and guide them so they can achieve their goals. One of the most important parts of our profession is to evaluate our students in this process and to be accountable for the evaluation so that the students can have a fair assessment of their current ability and set a goal for their further development.

If we take a look around us, many kinds of so called proficiency tests abound, starting from STEP, TOEFL, TOEIC, to name a few. We are seeing more and more ready-made exams on the market to 'assess students' real communicative ability'. Which test is the most useful and what kind of characteristics should the best test have? Are all those well-known tests perfect tests?

The purpose of this paper is to argue that there is no such thing as a perfect test that meets everyone's purpose. However, we can try to make a useful test. I will argue that in order to make a useful test, we need to carry out a careful needs analysis to make clear what it is that we want to measure. My purpose is not to offer a ready made formula as to how to design a perfect test, but to offer some guide lines as to how to make a useful test and what constitutes a test's usefulness.

Defining Usefulness

In this section I will introduce four qualities that would determine the overall usefulness of a test. As mentioned in the introduction, the fact that there are many kinds of tests on the market does not mean that these tests are the best tests that would meet our needs. We need to have a conceptualized framework to design a test that would meet our students' needs. Bachman (1996) defines a useful test as consisting of four pillars, reliability, construct ability, authenticity, interactiveness, impact and practicality. We will now look at each of the construct and discuss what each of the construct implies.

Reliability

The purpose of carrying out a test is to measure whatever needs to be measured and make inferences based on the measurement. In order for the test to be of use, this

measurement needs to be consistent. Reliability is defined as 'consistency of measurement.' (Bachman 1996) If a test was to be administered to a group of students on two different occasions, a reliable test needs to offer us similar scores on both occasions. If one student scores over 60 percent in one test and scores less than 40 percent on the other, the test can hardly be called reliable. There could be some reasons that could cause this fluctuation like the student's physical condition. We need to make every effort to minimize such risks.

Construct validity

Construct validity refers to the appropriateness of inferences we make of a given test. To what extent can we justify the inferences we make of the result of the test and to what extent can we generalize from the measurement? In designing a test we need to ask ourselves what we are testing and what we are trying to measure. In measuring students' listening comprehension ability, if the instructions of the tests are written in English using complex vocabulary, we cannot measure the construct we want to measure because the student might have a problem reading the instruction and do not understand the purpose of the test. Or just providing the students with multiple-choice questions to measure their writing ability would not yield comprehensive implication about their writing ability although the reliability of the test might be improved. So the issue of designing a test that is reliable and that has construct validity is a matter of balance. There is no designing a test that emphasizes one at the expense of the other.

Authenticity

Authenticity is another component of a useful test. Authenticity is defined as the extent to which tests tasks correspond to the target language use domain. For example, if we are designing a test to measure English proficiency of students who are preparing to study in American universities, the test tasks should include tasks that would correspond to what they would face when they arrive in American universities. Designing a test to measure their ability to answer a phone and take a memo as a secretary would not be authentic. So authenticity has to be defined with a view to students' needs and what they would need to do in real context. In a country like Japan, where students would not be used outside the classroom, it would be very difficult to define what the target language domain is, but at least we need to make an effort so that students can find meaning in sitting for a test.

Interactiveness

In sitting for a test as English learner, we utilize not only our linguistic ability but our knowledge about language, strategic competence, metacognitive strategies,

knowledge about the world or our affective schema. In designing a test for a junior high school student, a test whose topic is heavily academic would not be interactive because such a test would not engage the students' interests and developmental needs. However in designing a test that is interactive, we need to make sure that the test stays valid but ensuring that having knowledge of a particular topic would advantage certain students at the risk of disadvantaging others.

Impact

Bachman (1990) points out that 'tests are not developed and used in a value-free psychometric test tube, they are virtually always intended to serve the needs of an educational system or of society at large.' One of the important factors of any given test is that the test is carried out for the purpose of meeting someone's goal and there will be considerable impacts on those who are involved. For example, if a high-stake test like entrance examination starts administering listening comprehension tests, that would have impact on both the students and teachers. Teachers would have to accommodate their instruction so that students would not have problems later. For the impact to be positive students need to be informed about the test design and what it intends to measure and receive appropriate feedback. For teachers, impact of a test has great implication as well. As teachers we sometimes feel the need to teach to test rather than test to measure what we taught. We sometimes lose confidence to teach because of the disparity of what we believe we should teach and what we need to test. Although many are well aware of the communicative needs of our student, how many of us have wondered how to prepare our students for the exam and teach communicatively at the same time? One way to remedy this situation could be to change the way we test our students and bring instructional practice on line with current thinking in the field of language teaching. By this wash back effect, students might start to feel the communicative needs and start studying communicatively to sit for a communicative test. I would not go into what communicative means here. If interested, refer to Principle of Language Learning and Teaching by J.D. Brown, Prentice Hall)

Practicality

Finally practicality is the issue of resources both human and material. To administer a test we need to manage and allocate resources and decide whether the test is doable. Interviewing potential university students one by one would be interactive and maybe authentic and would have great wash back effects on teachers who refuse to use any English in English classrooms, but it is simply not practical.

Conclusion

As we have seen, to make a useful test, we need to define what the primary

purpose of carrying out the test is and discuss the characteristics of each test tasks. Each construct needs to be defined based on the characteristics of test takers and the testing environment. As English teachers we sometimes fall into self-denial and feel like looking for ready-made tests and materials that would give us some confidence so we can protect our dignity, but as professionals we need to be analytical and critical of the needs of our students and carry out clear needs analysis of what we are doing and why we are doing so in order to be accountable for our actions. In the end I would include guiding questions for making a useful test so that it would be of help in designing a test that is useful.

References

- Bachman, L.F. 1990. *Fundamental Consideration in Language Testing*. Oxford University Press.
- Bachman, L.F. 1996. *Language Testing in Practice*. Oxford University Press.
- Brown, H.D. 1994. *Principle of Language Learning and Teaching*. Prentice Hall
- Widdowson, H.G. 1978. *Teaching Language as Communication*. Oxford University Press.