
**INTERVIEW: A SPECIAL TALK
WITH PROFESSOR LYLE F. BACHMAN
CHAIR, UCLA DEPARTMENT OF APPLIED
LINGUISTICS AND TESL**

You are well known for your expertise in language assessment, and you have discussed both web-based assessment and paper and pencil assessment. What do you see as major developing trends in ELT language assessment?

⊙ One of the things is that there are quite a few studies now about computer- or web-based language assessment, not just research that has been comparing paper-pencil tests with computer-based tests. As people change from one format to the next, they are always concerned about whether the new test will work as well as the previous test for the same purposes. At UCLA for example, we have a paper-pencil ESL placement test which we use to place international students into ESL courses. For a number of reasons, we decided to develop a new web-based test. We now want to make sure that the new test will give us information that is useful for making placement decisions. Most of the studies that have compared paper-pencil to computer- or web-based tests have just looked at test performance comparing scores. Quite often they've found that their scores were the same or at least there was a high correlation. I think what they failed to recognize is that test takers may be processing the tasks very differently even though they achieve the same scores. So, in terms of test use, we might say we are happy to know that the web-based scores and the paper scores are giving us the same information. But at the same time, we have to wonder if the test-takers are processing the test differently and if the tests in fact are measuring the same thing. Those scores may

not mean the same thing even though they are ranking the students in the same way. This is of particular concern in classroom assessment, and I think one of the challenges of language assessment now is to reconsider the relationship between assessment and teaching. In a language course, assessment needs to serve the needs of learning. So, if teachers need to give tests to diagnose areas of strength and weakness or to give grades or to find out what students have learned, and if those programs are themselves computer-based and employ computer-assisted language learning, then the question is whether the assessments are aligned with courses and whether that alignment is multifaceted. In the past when the curriculum was paper-based with only textbooks and written materials and exercises, and the tests were paper and pencil, it was much less complicated. Now, we have multimedia classroom learning exercises delivered by the computer on the web and so forth. With multimedia tests, determining the relationship between all of those is a lot more complicated.

How does the movement towards web-based testing affect classroom teaching?

⊙ What we need to look at here is the idea of a washback effect, the idea that the test itself, the way we test, and the kinds of things that we test will have some impact on the way teachers teach, on the materials they use and the kinds of learning activities they will use, as well as what areas of language ability they will focus on. If you have a multiple choice test of grammar, then teachers may focus on having students memorize grammar points. And that's true whether we have a paper and pencil test or web-based test. In fact, it turns out that washback is a lot more complicated than that. It does not simply mean that all teachers teach exactly to prepare students for the test. While some teachers will teach for the test, others may continue to teach what they believe in anyway. So, investigating the impact of assessment on instruction is again also more complicated when we have a web-based test.

What do you think about web-based testing? Is it possible to use this technology when you are going to do performance-based assessment?

⊙ There are obviously some limitations on web-based technology. The main limitation right now is assessing what used to be called the productive skills of writing and speaking. So, even though we now have very sophisticated algorithms for scoring written essays that produce high correlations between machine scores and computer scores, we still do not know whether the computer is using the same criteria as the human scorers.

But scoring spoken responses is far over the horizon. We are not even close to being able to score anything other than matching a spoken response. There is a test which is available, called PhonePASS (www.ordinate.com) which uses speech recognition technology for scoring spoken responses. But those responses are very carefully limited. And they basically match the phonetic shapes, or sinewaves, of what test-takers say, with those in a data base of native speakers. So, PhonePASS can only assess what is in the database. We don't have a computer yet which can interact with human beings the way two humans can interact. Second, we don't have sophisticated enough speech recognition programs to score open-ended speaking. I think that is a major limitation right now that will prevent us from using the web for anything other than capturing spoken responses. We can give a computer- or web-based test in which students can speak their responses into a digital recorder and then that can be scored by a human being. But, it's not interactive, and it's not like a conversation or interview or anything like that. I think right now the technology will continue to place some limitations on what we can do. On the other hand, the kind of things that we've been doing in our webLASS project is to try to move a little bit towards language use, where we have limited production responses. Students have completion tasks where they actually have to write short answers, and we develop a scoring algorithm to score those. So, that is a small step towards full written production, a small step away from multiple choice. Students have to produce language. That is something that we can do now with a web-based test that was not feasible before. In some

ways, I think the technology enables us to move a little bit more towards performance assessment. But there are still some limitations on that. Certain kinds of performance assessment will be very difficult to at least score by computer. We can deliver them by a computer but it will be difficult to score them.

Do students develop negative attitudes towards the use of computers because they can't use familiar paper and pencil strategies such as circling?

⊙ In general, we've found that students react very positively to web-based tests. There are certain aspects, of course, to which students react negatively. One example is their desire to interact with a reading passage in a test. In a paper and pencil test they can underline, write comments or translations of words in the margin, and so forth. In a web-based test we can let students do some of this. The technology is there. We can modify the tasks. We can use software to let students highlight, but I don't know whether we are able to let them write in translations or not. They can highlight, and they can circle. We can do that. We learn something from that. If you give students a reading passage and we don't allow them to interact with it the way they do with paper and pencil forms, then they will not be able to use some of those strategies. That raises the questions, "Why do we want them to do that?" or "Why do we want to prevent them from doing that?" If we don't allow them to underline and circle, are we measuring something different from if we let them underline and circle? These are difficult questions to answer.

What do you think is the most important assessment challenge for teachers at the moment?

⊙ I think, the most important assessment challenge, with respect to teachers, is that they are still not adequately trained to do assessment. In most countries, teachers can become certified language teachers without taking a single course in language testing. But, on the one hand, we are saying that we would like to have more classroom assessment to aid learning and to collect

feedback. In some situations, school districts and school systems measure the effectiveness of teachers by test scores. So, you have large scale standardized assessments that are used. If students get high scores, the school gets rewarded or the teachers get rewarded. However, we would like to be able to base those decisions not just on standardized tests but also on classroom assessment. So, we'd like the teachers to be involved in that because teachers know students the best. On the other hand, some people said we can't trust the teachers because they don't know how to assess. They can't do it reliably. They can't do it validly. The question is if the teachers can't do it, "Why can't they do it?" Because all their courses, all their experiences, and all the training they've had have been focused on how to teach, how to create learning activities, how to assess the needs of their students informally, how to tailor the textbook to the students' needs and so forth. But they never learned how to assess. And in fact they always have to give tests. I've never worked with a group of language teachers who said they didn't have to do assessment of their students' performance and ability. So I believe that the adequate training of classroom teachers in language assessment is one of the most pressing challenges facing teacher educators today.

What is the significance of university exit exams in English? Do universities need to give students exit exams to validate what they have been teaching to the students?

⊙ That is something that has become very popular around the world, not just in Thailand. There have been benchmarking or exit exams in the United States and Canada. I have been involved in helping develop one in Hong Kong. The rationale often arises out of dissatisfaction, on the part of potential employers, with the level of English that universities are producing. It seems to me that the real solution is to look at the way English is taught at university. We need to try to find out why the language programs at the universities are not successfully reaching the goal of producing students who are functionally proficient in English. And then try to figure out what can be done. When an exit test is given, it typically means someone is looking for an easy solution. If you have an exit

test, think about what the possible outcomes might be. Ideally, if language instruction is successful, everyone would pass an exit exam, OK. However, if that happened then people would say the exit exam was too easy. They would question whether it was really giving us the “truth.” The alternative would be that everybody fails it, so then you have to have a system that lets people take it more than once, several times. Another problem with the exit exam is “What is the domain to which you want the exit exam to generalize?” At Chula, for example, some of your graduates will go on to graduate schools here in Thailand. Others will go to the United States, or Australia. Some may go to Japan. They will go to different places. Some may enter the workforce. So, there are lots of different places where they will be using English after they finish their college degrees. So, what do you use as the content of the test? Is it going to be a business English test? Is it going to be English for graduate school? English for engineering? English for communications? English for finance? What kind of a test is it going to be? Or is it going to be based on where they have been, that is, based on classroom assessment? So again, the question is if we want to use classroom assessment as part of the summative assessment, then we need to be able to count on the teachers’ assessment being reliable, valid, and so forth. So, I think the biggest need in teacher education is better training in assessment.

In your opinion, what seems to be lacking among language teachers’ perception of the role of assessment?

⊙ I think teachers often have the idea that teaching is here and that assessment is there, and ne’r the twain shall meet. They feel that the experts kind of know how to do the testing, right? Teachers often don’t see the connection between the curriculum, teaching, and assessment. I think it is important for language testers to work more towards helping teachers understand that assessment is really a way of collecting information, of collecting feedback. Teachers need to realize that testing is not just something that is done at the end of a course by experts.

Can we say that assessment gives feedback on teacher performance as much as student performance?

⊙ How often do teachers give a test at the end of a year of instruction and find their students not doing very well? In such a case you have to ask, "Why did the students not do very well?" It might be that it was not a good test. Maybe the test was too difficult, you could say. But we should not dismiss the possibility that students did badly because the teaching was not very effective. And if teachers are insecure about their ability to give good assessments, the tendency is to say, "Well, that wasn't a very good test," rather than, "Well, this was a true reflection of my teaching." So, if my students get some things right, but they all get some things wrong, this tells me that these are areas where I have not been very effective as a teacher.

Can you share your views on summative and formative forms of assessment?

⊙ I think formative assessment can be just as important as summative. If you give a diagnostic test or if you assess your students at the beginning of a course to find out what their areas of strength and weakness are, and if you don't assess that accurately or if you make errors in judgment, you can waste students' time by asking them to focus on things they already know and by ignoring things they don't know. So, even for formative assessment, it is important for teachers to understand how to test. What kinds of tasks are appropriate? What qualities do you need to think about when you assess? What inference are you going to make? And this is something teachers are not trained to do. For summative assessment, I think it is important that we use multiple sources of information. When we need to make large scale summative decisions about school placement, about curricula or about individual students, or when we are giving grades at the end of a course, we might use a standardized test as part of the information. However, it would be better if we could use additional sources of information as well. So, okay, these are the courses that we have used to teach English. Let's make it a classroom achievement test.

So, it's a very very complicated issue. There's no easy solution. Just giving a single test will not solve the problem.

What should teachers do when faced with such assessment decisions?

⊙ I think often we are given a mandate to develop a test by authorities, and as a professional you have to sometimes ask yourself, "Do I want to get involved in this?" "Is this professional?" "Is it ethical?" "Is it in the best interests of our students?" "Is it in the best interests of my institution?" And you have to ask yourself whether it is what you want to do. Sometimes you are in situations where you don't have a choice.

Any suggestions on areas of research that haven't been touched upon enough?

⊙ I think things like performance assessments or authentic assessments are going to be around for a while. And I think the profession is moving in that direction. The reason I say this is because this is happening not only in language testing but also in educational measurement. And so people are beginning to develop measurement methods that are sophisticated enough to deal with performance assessments. There are still lots of problems to be solved with performance assessments. But I don't see us going backward now and saying that performance assessments are too difficult to do, or that we can't do performance assessment any more, or that we have to go back to multiple choice tests. If we do move away from performance assessment, it will be somewhere else, I think.

Is there no longer any use for multiple choice tests?

⊙ I wouldn't say that you should never use a multiple choice test. There may be some situations where a multiple choice test is the most useful way to go. If you have lots and lots of very clear discrete learning objectives in a language course for example, a multiple choice test might be fine. One of the advantages of a

multiple choice test is that it is very efficient, so that you can collect a lot of information in a very short period of time. Thus, if you want to assess a large domain, a multiple choice test might be useful for that. The question in my mind is, of course, whether that domain really consists simply of bits and pieces of language or whether we really want to assess language use. A linguistic analysis could be used to identify the bits and pieces here. We could analyze the structures we would need to use language. We could analyze the vocabulary. We could analyze the cohesive devices. And at the end we'd come up with a list of all the pieces of language that we've used in our conversation. But we still would not capture the kinds of interactions and the functions that we would perform. And so the question is whether or not we can in fact measure language use just by measuring bits and pieces. So, my advice would be to try something else first. We should try to get students to work with a task where they have to actually produce some language. And even if it's just a short answer, it is a small but, I think, important, step.

Do you believe web-based assessment readily yields itself as a perfect tool here?

⊙ I would be the last person to claim that web-based assessment is perfect. However, there is one great advantage of web technology over either paper and pencil or computer-based assessments. The system that we've developed at UCLA can be used anywhere in the world anytime. If students want to take a self-assessment of their own English for whatever purpose, it is a low-stakes test and there is no reason for them to cheat or to ask their friends to take the test for them, they can do that now, in their home, their classroom, or wherever. And potentially what that means is a lot of that sort of assessment could be done in a way that doesn't take away from classroom time, but which is still integrated with or articulates with the content of courses. And that's something you can't do with a computer-based test. Because a computer-based test is basically locked into a computer, you have to have something in the hard drive that compiles data or you need to have a CD. But the beauty of the web is that anybody can log on, as long as they have a web

connection and a password. I think this is the beauty of web-based assessment.

