

Interview with Antony John Kunnan on Language Assessment

Interviewed by
Vahid Nimehchisalem

Faculty of Modern Languages and Communication, Universiti Putra Malaysia
E-mail: vahid@upm.edu.my

Received: -
doi:10.7575/aiac.ijels.v.3n.4p.56

Accepted: -
URL: <http://dx.doi.org/10.7575/aiac.ijels.v.3n.4p.56>

Published: 31-10-2015

Introduction

Antony John Kunnan is a language assessment specialist. His research interests are fairness of tests and testing practice, assessment literacy, research methods and statistics, ethics and standards, and language assessment policy. His most recent publications include a four-volume edited collection of 140 chapters titled *The Companion to Language Assessment* (Wiley, 2014), a four-volume set of 75 previously published papers titled *Language Testing and Assessment* (Routledge, 2014) and *Talking about Language Assessment* (Routledge, 2015). He was the founding editor of *Language Assessment Quarterly* (2003-2013) and the founding president of the Asian Association for Language Assessment. He has held academic and professorial appointments at universities in Bangalore, Ann Arbor, Los Angeles, Taichung, Hong Kong and Singapore. Currently, he is a research consultant at the Center for Linguistics and Applied Linguistics at Guangdong University of Foreign Studies in Guangzhou.

The interview you are about to read here was recorded before Kunnan's plenary talk at the 4th International Conference on English Language Teaching (ICELT2015) in Melaka, Malaysia. My colleague, Dr. Seyed Ali Rezvani Kalajahi, managing editor of the *International Journal of Applied Linguistics and English Literature (IJALEL)* also accompanied us during the interview.

The interview

VN¹: What is your definition of language assessment?

AJK²: Language assessment is a system by which we can judge someone's ability in a language. It could be a single test. It could be multiple tests. It could be an observation. It could be what we typically know as a standardized test. It could also be a non-standardized test and if you look at the practice of language assessment, it is done every day in every school or college in the world. But most of the time we sort of tend to ignore this and only pay attention to the big standardized tests such as TOEFL or IELTS, GRE and so forth. So, it is an everyday happening. This is the first point and with from this point of view, language assessment is ubiquitous.

The second point is that it is social practice like education. We all believe that it's useful to go to school or have home schooling. Very few societies will argue that children do not need to go to school or have schooling and assessment in its broadest sense in school context, primary school, high school or even college is supposed to help the learner to understand his or her strength and weaknesses. It also helps the teacher or the instructor to know on average how everybody is doing in his or her class, know their strengths or weaknesses so that he or she as the teacher can modify, change, reconstruct, or re-manage the teaching instruction process. So, those are the main goals. But, today we have other kinds of uses of language assessment such as assessments in the work place. Imagine a company or a bank wants to hire someone and there are many applicants; how do they make a decision to hire? Similarly, there are other kinds of jobs, such as nurses, doctors, hospital administrators and secretaries at the front desk who need to talk to their patients. So, if you want to hire a bilingual speaker, for example in California, and if the applicant says, "I can speak English and I can speak Spanish at a high level," how do you know if this true? Is it enough to just take her word? Or do you want to have some process by which you can make the judgment?

Another workplace context is the critical exam for air traffic controllers. Air traffic controllers all over the world are supposed to speak English so when a plane goes from one jurisdiction to another, the advice given by air traffic controllers to pilots has to be continuously in English, which means air traffic controllers and the pilots and copilots need some specific English language training. So, ICAO, an organization in Montreal, which governs civil aviation for the whole world, has a program for assessing English language. The assessment certifies air traffic controllers in terms of their ability to speak English.

And, in the past ten years, language assessment it has also been used for immigration and citizenship. So, if you look at Europe, many countries have introduced legislation to start language tests that persons seeking immigration or citizenship seeker needs to pass. They are already in place in the U.S., Australia, and the UK. Now you also have language assessments in Netherlands, Korea and Japan with Spain, Italy and other European countries are considering

¹ Vahid Nimehchisalem

² Antony John Kunnan

language assessments. So, many countries are instituting language assessment procedures to determine whether immigrants or citizenship seekers are able to speak that dominant language of the country.

VN: The way you look at language assessment has a fascinatingly wide scope. You are saying that language assessment is not limited to universities or schools but it could happen even when a mother is speaking to her child; every minute she is assessing her child's language.

AJK: Yes, parents always give advice to children to be polite or to be asking questions politely, so they are also part of teaching-learning process. Also, whatever we are teaching and students are learning, we like to assess to find out whether everything is ok.

VN: Exactly, and that's what takes us to the next question which is how would you complete the sentence *A world without tests is ...* with a metaphor?

AJK: A world without tests specifically maybe pointing towards formal and standardized tests but if you look at testing in a broad sense, like I did, it is there everywhere. So, a world without tests is like there is no education or teaching and learning.

VN: So a world without test is a world without education.

AJK: That's right. Of course, we are not thinking of tests in a narrow sense.

VN: How does language assessment differ today from what it used to be?

AJK: If you go back about 50 or 100 years, testing was done way narrowly. It was mainly in the school context. People went to school and they studied some aspects of language, such as grammar and vocabulary, spelling, punctuation, diction maybe writing answers to questions, things like that, or studying poetry, plays, fiction, and so on. Some of the assessments we see from 50 or 100 years ago might look odd to us today. I have some examples in my presentation tomorrow, which will make us wonder if I took this test today, I might fail. There are questions like:

- *What are the nine rules for capitalization?*
- *Parse the following sentence!*
- *Draw a tree diagram!*

And, this is for the 8th Grade. I am going to show you examples of the 8th Grade test in Kansas in 1895. Language at that time was taught as a body of knowledge like physics, biology, chemistry, or history. We had to study it, memorize it, and repeat or regurgitate. So, from that model we have moved constantly in the last 50 years to more towards language for communication, to communicate our ideas. Because of the focus on communication, many of the assessments you see today are also focused on communication. So, we have, for example, listening and speaking tests which did not exist about 50 years ago or 100 years ago. Not just for practical reasons but also because the philosophy was that you needed to know rules of grammar, vocabulary, conjugation, tense and aspect. Those were the highlights of school rules and colleges that had instruction in languages. But, today that is not the focus; people are studying languages for communication. So, whether you learn your first, second, third or fourth language, whether you learn it as a child, whether you learn it as an adult, the focus has changed. Because the focus of instruction has changed, so has the focus of assessment. So, today we have more assessments that are focused on communication and on listening and speaking. We have come a long way.

VN: Great, it's like reading a book as you are talking. My next question is something related to the previous question. What do you think language assessment will be like in the years to come?

AJK: As you know, we are in this technological age. Technology is affecting us: we use cellphones, for example. We could not have imagined this 20 years ago. So, technology is definitely going to affect assessment. Specifically, let's take tests which use computers. Today there are more and more testing agencies that are interested in presenting clips, photographs, or even short videos for people to watch and then to comment on or respond to. These sorts of tasks were not possible 20 years ago because of the bandwidth, memory, and all those limitations. The biggest change, however, is going to be in one other area, automated scoring, automated scoring of writing and automated scoring of speaking. You might be familiar with how sometimes when we call a bank, a voice says, "Please say your name!" or "Please say your password! And, we say it, and the system recognizes it. What is happening here is speech recognition technology recognizing our voice. Similarly, in a more complex way, there is a phone-based test called *Versant*, which is administered through a phone. You call on the phone and they ask you a bunch of questions and you respond to the questions and they assess your speaking ability. Some of us may have questions about some of the tasks in many of these tests because they may not resemble natural listening-speaking. While automated scoring of speaking is emerging as an important sort of innovation in the field, a more widely used innovation is automated scoring of writing, especially of short essays, based on some text that has been given or based on a prompt that requires a somewhat narrow response. Automated scoring may not do as well if you ask someone to write a story but if you are given a topic like *Discuss the haze in Malaysia and its causes and possible solutions*, you can actually train the computer software through previously written essays of high quality and then look at some critical linguistic features like essay length, lexical diversity, topic-specific vocabulary, etc. There are three agencies that are leading the charge: *MyAccess* developed by Vantage Learning, *WriteToLearn* by Pearson, and *E-Rater* by ETS, Princeton. *MyAccess* and

WriteToLearn are available for purchase for use in schools and colleges around the world where English is used. They claim that they do a very good job essay scoring and feedback that they have trained the software for. So, if you are a school or college teacher sitting somewhere in, say, Malaysia, Indonesia, or China, if the school buys the product, you can assign the students one of those essays trained by the software. The software will give you an essay score within a few seconds along with feedback.

VN: And, you do not have to worry about the marking.

AJK: I presented two papers recently based on two studies, one on *MyAccess* and one on *WriteToLearn both with my former MA students*. It was shown that the comparison of scores between human raters and automated raters is getting better. The correlations are about 0.70. However, what is not getting better is when you compare teacher feedback with automated feedback in terms of errors in the essays. So, when a student writes and makes errors, in many areas, the total number of errors detected by the teacher, say, if it is 100, the automated scoring is only able to identify about 60% of the 100 errors, sometimes 70% but not more. Furthermore, there are some areas with which automated scoring is very good at: spelling, punctuation, mechanics, but not at all good at word choice, coherence, cohesion voice, etc. So, this is good news in a way. Good news from the point of view of getting a quick score but not very good news in terms of feedback. I think at present, automated scoring of writing can be used by schools but only in tandem with teachers. The writing teacher cannot avoid scoring essays but the automated scoring system can capture all the superficial, easy-to-spot errors. So, if a student makes 100 errors in an essay, and if the computer software can identify 50 or 60, that's very good because the teacher can then focus on the other 30 or 40. The teacher does not have to look for spelling or punctuation and all of those errors which if some software can do, so, why not let it do it? So, what I think we have to at this point is innovation in terms of automated scoring in writing.

VN: You answered my next question, the fate of high-stakes exams and high-stakes international proficiency tests.

AJK: There are more agencies now than before administering regional, national and international assessments. Language assessment has become a multi-billion dollars business. It wasn't the case before, but now with the explosion of English language teaching and learning, around the world for businesses, offices, universities, and colleges.

VN: And other languages like Chinese.

AJK: But, even in China, there is a big push to learn English, so local schools in small towns want to teach English. If they find out a local school is not teaching English, parents are going to send their children to another school. So, English language assessment has become big business. And, because it's become big business, more and more players are getting into it. For example, the British Council is now involved with *Aptis* which is an assessment of teachers' English. If a teacher says I can speak English, well, there would be a test and I think the Malaysian government is interested in adopting *Aptis* or some version of it to certify persons who are capable of teaching English. This is a new area that has involved language testers.

VN: So, you believe that language assessment in its high-stakes form will not go extinct but it may change and evolve through technology.

AJK: Yes, because of the simple issue of supply and demand, demand is very high for places in schools, colleges, universities, the workplace, and immigration and citizenship. People want to have some form of assessment and if language is involved, they will have a language assessment.

VN: At the beginning of the interview, you mentioned a question from 100 years ago that you are going to present tomorrow, and I think if we do this and we keep on doing this, people, Antony Kunnan of a hundred years later in their conferences, will present our questions and people will be laughing at our multiple-choice items. As you said because it's a business, people forget that they are dealing with language and people. My next question is about validity, what does validity mean to Professor Kunnan?

AJK: I have edited two volumes, one on validation, one on fairness. But, in one word, it means meaningfulness. Is the assessment meaningful? So, for example, if you say I am going to design a test of reading ability in English or any language, the question is, is the test meaningful in terms of what it's supposed to assess. If we say reading, is it assessing reading? What type of reading? What is the construct of reading? Validity has a long history. There have been many phases of interpretations of validity. Today there is less talk about validity and more about validation and currently people are interested in developing validation arguments for an assessment, to support on assessment. Therefore, rather than talking about the validity of an assessment, the current approach is to develop arguments that can be presented to the public regarding an assessment.

VN: I usually tell my students when you want to define validity in a word remember relevance, but I think, I will go for meaningfulness from now on. The next question is Why do you think in some testing systems, it's common to assign a larger range especially for average levels like in IELTS if out of 40 questions you answer correctly to 16 to 25 questions correctly (a range of 9 points), your converted band score will range between 5 and 6, but if you want a band score of 8 to 9, you must answer 35 to 40 questions correctly (a range of 5 points). As we see, the range is unequal between intermediate and advanced levels. Is it fair?

AJK: One of the ways to study this issue is to actually take many essays from students and look at the linguistic properties of the students' writing or answers and then examine the clusters of papers to see whether there is a real

difference between what is an 8 point paper and what is a 7.5 point paper and the 7 paper. Then, you would be able to justify having a 7 and a 7.5 and an 8 and 8.5 score points. I am assuming that an organization that runs IELTS has done this kind of research and they have noticed significant difference in the quality of essays or quality of writing.

SARK³: If you have noticed in speaking and writing section of the IELTS, scores such as 7.5, 8, 8.5, 9 are meaningful yet when it comes to test of listening and test of reading it likely to be meaningless. In fact, there is a noticeable difference while scoring receptive and productive skills. For instance, when I compare a writing, which has got 7 and 8, sometimes I can spot kind of differences but when it comes to listening and reading, it's a bit challenging and I am unable to see the rationale behind it. So, a candidate has to answer 39 or 40 to be able to get 9, whereas if one answers 35 to 36 questions correctly in the listening test, the score would be 8. 32 to 34 correct answers in the same test would be 7.5. So, this is where the actual difference emerges.

AJK: I think these cut scores at different levels have to be defended by doing some psychometric and substantive analysis; otherwise as you say, it puzzles everybody and for the teachers and for test takers, there should be more transparency. I would suggest looking for a research report. If they have not done it, we would be happy to see a research report that justifies their decision making in terms of cut scores for different bands.

VN: That's it, you just answered my last question.

AJK: Thank you. If you have any questions, you can send the questions to my email (akunnan@gmail.com).

VN: Sure, any last words?

AJK: Last words, become a language assessment professional so that you can change what you don't like. There's no point complaining about language assessment matters from the outside; you've got to be in it. It's always easier to complain from the outside.

³ Seyed Ali Rezvani Kalajahi