LTJ 28 2

Interviewer: From the University of Leicester in the United Kingdom. This is Glenn Fulcher with another issue of Language Testing Bytes.

Recently, the Journal has seen a steady stream of papers on vocabulary testing, and in this issue 28(2) we have another one from Marta Fairclough on the use of a lexical recognition test using the yes/no format for Spanish/English bilinguals. We thought that it was time to consider just why vocabulary testing is attracting so much attention at the present time.

We're very lucky in this podcast to welcome Professor John Reid of the University of Auckland in New Zealand, who has agreed to provide us with an overview. Professor Reid, many of you will know, was the author of *Assessing Vocabulary*, in the Cambridge University Press Assessment Series, which was published in 2000.

The recording of this podcast took place in New Zealand and, as I couldn't get there as much as I would have loved to have gone, we are grateful to Rosemary Vet [SL 01:15] for asking the questions on my behalf. We hope you enjoy this issue of Language Testing Bytes.

RV: Welcome to Language Testing Bytes, John, and thank you for agreeing to talk to us about vocabulary testing.

Respondent: Well, thanks for the opportunity; it's been one of my prime academic interests for some time now.

RV: First of all, can you talk about the role of vocabulary testing within the field of language assessment, it seems to have become almost a sub-discipline, is it that vocabulary is somehow different from other aspects of language knowledge and use?

Respondent: Certainly, when I became involved in vocabulary testing in the 1980s, it was fair to say that vocabulary testing was outside of the mainstream of the field. Up until that time, say up until the 70s at least, vocab tests had a central place, especially in American psychometric testing, so multiple choice items to test knowledge of particular words were quite a routine component of tests like TOEFL, for example, and in the TOEFL case that applied right up until the 1990s.

But, in the 70s and 80s communicative approaches to language teaching started to have quite an impact on testing as well, and so the design of language tests moved away from individual test items to more task-based test design, and so vocab tests of the traditional kind were no longer considered to be valid measures of proficiency or achievement.

So, that meant that, at best, there was a need to reconceptualise vocabulary assessment to fit into a task-based testing framework, and the place that it found, essentially in speaking and writing tasks, was as one of the criteria that raters use to evaluate learner performance, so they look at vocabulary use or range of expression, and maybe give an analytic rating for that category along with various others.

Now, this is what I call embedded vocabulary assessment where you're looking at vocabulary within the context of a larger construct, so it's like speaking proficiency, or academic language ability, in a more general sense. And a lot of language testing researchers weren't particularly interested in vocabulary testing but I think we had to ask, is there still a role for the kind of discrete assessment of vocabulary knowledge that had traditionally been so important? I think it still doesn't really have much of a place in proficiency or achievement assessment but there are other purposes of assessment as well, placing students in

language classes, measuring progress in the school or the classroom, and diagnosis, which is becoming quite a trendy area in language assessment. I think, for these purposes, vocabulary tests are still very relevant and useful.

RV: I see. That leads us on to looking at particular vocabulary tests that have become popular, particularly for placement and diagnostic purposes. For example, recent publications on vocab testing seem to have focused on the assessment of vocabulary size, especially by using the yes/no format. Why do you think it's become such a popular tool?

Respondent: Well, first I could say, if we look at vocab size testing generally, there's a long history of research in this area particularly by L1 reading researchers, who have long recognised the importance of the vocabulary knowledge for fluent reading comprehension. The basic procedure when we set up a measure of vocabulary size is to take a sample or words from a word frequency list and then we'd devise a test to see whether each of those words is known or not, and then based on the test scores we make an estimate of the total vocabulary size of the native speaker or the second language learner.

Now, to make a good estimate you preferably need quite a large sample of words and the trade-off of that is a simple assessment task. So, that's the attraction of the yes/no format, it's really the simplest kind of test task where you present a series of words and you ask the test takers, do you know this word or not?

So, it's really a self-assessment task rather than a test per se, and the way that we check that the learners are not over-stating their knowledge is that, as part of the task, we include a certain percentage, typically say 20-40% of non-words, words that are not real words in the language, so that if the learners claim to know some of those words it provides a basis for adjusting the estimate of their vocab size downwards. And, there have been some quite sophisticated scoring formulas that have been proposed based on certain assumptions about the typical response behaviour of learners in undertaking this sort of task.

The research shows that the yes/no format is more effective in some contexts than others. Researchers in Japan, for example, have found good results with Japanese learners of English, but, on the other hand, there are a series of studies in Belgium, which showed that the same task was actually quite problematic when it was used with Dutch speaking learners of French.

So, we can't always assume that it will be a workable measure, that learners will respond to the task in the way we'd like them to, and, since it is a self-assessment task, it's not suitable for use in high stakes assessment situations, but for those sort of purposes I mentioned earlier, for placement, for diagnosis, and for certain research purposes, for example, in getting a quick and easy measure of broad proficiency level, the yes/no formula seems to work effectively enough in those kind of situations.

RV: Okay, I think that that really helps to explain why the yes/no format has become so widely used. Can you briefly talk about other vocabulary tests that have been developed in recent years, and explain what aspects of vocabulary knowledge or use they're designed to assess?

Respondent: I think the first point I'd want to make is most discrete vocabulary tests, as I've defined them, where you're testing knowledge of individual words, these tests tend to rely on a fairly limited number of formats. I mentioned yes/no just now, and also multiple choice in the context of TOEFL and other traditional psychometric tests, so those are still very commonly used formats.

Another one is a matching task where you present learners with groups of words and groups of definitions and they have to match the words with the definitions. A fourth format is gap filling where learners are required to recall

the meaning or the form of the word rather than recognising it. And then there's translation which may be a part of some of the formats I've already mentioned but can also be seen as a testing task in its own right.

So, even today, I think those five formats tend to account for most vocabulary tests that are commonly used. The innovative vocabulary measures tend to be used in research studies rather than operational tests. One reason is that the conventional formats lend themselves well to computer administration, which is becoming more important these days in testing practice.

One important new measure I can mention is Paul Nation's vocabulary size test, I've talked about how the concept of vocabulary size measurement, he's now devised a test that can be used with both native speakers and second language learners. It's based on a word frequency data from the British National Corpus. Interestingly he's chosen to use the multiple choice format rather than yes/no or matching, because he considers it's important to verify that learners do know the words. There's actually a good validation study of this test by David Bigler, which appeared in *Language Testing* in 2010.

As a vocab size measure the VST, the Vocab Size Test, essentially replaces the vocabulary levels test, which I guess is one of the most famous vocabulary tests in language teaching, also devised by Paul Nation, that wasn't really a vocab size test at all, although it's often treated as being one.

I'd like to also move on and talk a bit about the concept of depth of vocabulary knowledge, going beyond just testing learners' knowledge of word meaning to assess other aspects of word knowledge. There are a couple of depth measures that have become quite widely known, one of the them is my own word associates format where the test takers are presented with a target word and then a series of other words which are related, or may be related, to the type of word in various ways. So, for example, if they're given the target word, edit, there will be other words like, revise, text, and publishing, that are related to edit in various ways, and they are to distinguish those words from ones like, pole, risk, or surface, which are not related to edit.

So, it's based on the concept that part of our knowledge of vocabulary is understanding which other words are related to a particular word. This format has been tried out in a variety of contexts with some encouraging results, but, as far as I know, it hasn't been adopted for operational testing purposes.

Another measure is the vocabulary knowledge scale, which was also designed to measure depth of knowledge. It was developed as a research instrument by two Canadian scholars, Sima Paribakht and Murray Fisher. Like the yes/no format it involves self-assessment in the sense that for the first three steps in the scale learners rate how well they know the meaning of each of the target words that they're presented with, but as they go to the higher levels of the scale they are also required to give a synonym or translation of the word, and also use the word in the sentence.

Again, the VKS has worked well as a research tool for Paribakht and Fisher, and for other researchers, it's also been criticised on various grounds, both conceptual and technical. So, there are a number of interesting innovations like this but they've tended very much to be used for research purposes, and so in practical assessment contexts it's really those five basic formats that predominate.

RV: Okay, perhaps we can finish up by talking a little bit about the challenges that are faced by researchers who are developing vocabulary tests and conducting validation

studies with vocabulary tests, what do you think are the major challenges facing the field?

Respondent: Well, again, maybe giving a little historical perspective, I think there was a problem with some widely known vocab tests that were developed in the 1980s and 90s, the authors… because language testers weren't really interested in vocabulary testing the people who wrote these tests were vocabulary specialists who designed the measures for their own purposes, and didn't necessarily have a good background in language testing. So, even when these tests were published in research articles there was inadequate evidence for the technical quality or the validity of the tests.

So, I think the situation has changed a lot since that time, some of those tests I've referred to have subsequently been subject to critical scrutiny, and, in some cases, revalidation. If we look at journals, like *Language Testing* and other major journals in our field, there are articles, good quality articles on vocabulary assessment appearing quite regularly.

Also, another development quite recently is the publication of a couple of research manuals on vocabulary, one that's already appeared by Norbert Schmidt, and another one coming out written by Paul Nation and Stuart Webb. So, I think these will have a positive effect, but it's certainly one challenge to ensure that vocabulary tests meet high quality technical standards.

The other kind of challenge that I want to mention is the need for us to go beyond testing knowledge of individual words to ask the question, how do we assess knowledge and ability to use multi-word lexical units, and that's a cover term for a whole lot of different things, like phrasal verbs, for example, to put up with something, or colocations, like we say, heavy rain, but not strong rain, or thick rain. And, idioms like, to shoot the breeze, formulaic sequences as when the chair of a meeting says, I declare the meeting closed, and lexical phrases are another category, like, as a matter of fact. So, more and more, particularly under the influence of Corpus analysis were recognising just how important part of everyday language use is represented by these kind of multi-word units.

There's been a little bit of work by researchers looking at how to assess knowledge of collocations, but more broadly the concept of formulaic language, or whatever you want to call it, is much more comprehensive and much more pervasive, so we need to ask the question, to what extent can we use discrete test items to assess these lexical units as if they were the same as individual words? Or, do, inevitably, they have to be assessed more integratively through task-based formats, because I think context of use tends to be quite important for a lot of these multi-word units.

So, this is an area of rapid growth in Corpus Linguistics and vocabulary studies generally, but we've only just started to think about the implications for vocabulary assessment.

RV: Well, John, you've provided us with some really good insights into current practices and challenges of vocabulary testing. Thanks for joining us on Language Testing Bytes and sharing your views with us.

Respondent: It's my pleasure.

Interviewer: Thank you for listening to this issue of Language Testing Bytes.

Language Testing Bytes is a production of the *Journal of Language Testing* from Sage Publications.

You can subscribe to Language Testing Bytes through iTunes, or you can download future issues from ltj.sagepub.com, or from languagetesting.info.

So, until next time we hope you enjoy the current issue of *Language Testing*.

[End of recorded material]
NOTES: [SL 00:00] Sounds Like