

L TJ_29_4_Yes_No_Vocabulary_Tests

[Start of recorded material]

Interviewer: From the University of Leicester in the United Kingdom, this is Glenn Fulcher with another issue of Language Testing Bites. Issue 29.4 of Language Testing sees another paper on vocabulary testing, this time taking a look at the ever popular Yes, No test. The authors are Anna [unintelligible 00:00:26] and Norbert Schmidt, both from the School of English Studies at the University of Nottingham in the UK. We invited them to join us on Language Testing Bites to discuss the scoring issues that they have been investigating. Welcome to Language Testing Bites to talk about your research into Yes, No vocabulary tests, which is published in this issue of the journal.

Respondent 1: Thank you very much for having us here today.

Respondent 2: We are really excited to talk about vocabulary testing with the larger testing community.

Interviewer: There has been a growing interest in vocabulary testing over recent years, and last year John Reid gave us an overview of current research in issue 5 of the podcast. But you have kindly agreed to talk to us about your research into the most popular kind of vocabulary test. To start off, can you explain for the listeners just what a Yes, No vocabulary test is?

Respondent 2: Yes, No tests are one of the most commonly used type of vocabulary tests, and really they are one of the simplest formats. Basically the test writer samples words from the target population that you want to test, and then you place those target words on a piece of paper in lists, and then the testers simply check whether they know those words or not. So essentially a Yes, No test is a yes, no decision. Yes I know this word, or no I don't. The other common name for this type of format is called a Checklist Test, because the tester simply checks the words on the list, whether they know them or not. What does it mean to know the word though? That could range anywhere from just being able to recognise the word, all the way up to being able to produce and use the word appropriately in any context you want. So I think that the instruction is really important. It is probably best to use Yes, No tests as receptive tests, that is where the testers look at the words on the checklist, recognise the form, and then are able to recall the meaning, so the type of terminology I like to use it would be a meaning recall level of knowledge. Now Yes, No or Checklist Test has some advantages of course, otherwise they wouldn't be used. The main advantage is that they are really quick. So you can have a lot of target items on the test and have a relatively high sample rate. Of course it has a converse side, so there are disadvantages as well, and the main disadvantage is that there is no demonstration of knowledge, and what this means is that learners can overestimate their vocabulary knowledge on the test, that is they can say that they know words that they really don't.

Interviewer: Now, as you have indicated there, a potential problem is that a test taker will just say that they know the words, which would lead to an overestimation of their vocabulary size. Can you say what the traditional approach to solving this problem is and why scoring these tests is so problematic?

Respondent 2: I said that Checklist, or Yes, No tests are some of the simplest formats, but that is not exactly true, they are very simple to write but in fact they are quite complicated to score. So you mention this problem of overestimation. The main way we adjust for overestimation is to include non-words, so these are word forms that are legally possible in a language, so in English a non-word could be something like float, it is perfectly possible but it doesn't happen to be a word in English. And typically we would put 25 to 33% of the items on a test would be non-words. Now the assumption is that if testers check any of these non-words, of course they don't know those words because they don't exist, and so they are not really being careful enough, and so we can assume that some of the real words that they are checking may not be known either. The problem is how do we actually use these non-words to adjust the scores down? There is two main ways, the first one is to delete any data that has over a certain number, or over a certain

percentage of non-words checked. So I use this methodology with my co-authors in a recent modern language journal article, in that research we had a Yes, No test with 30 non-words. And so we only accepted the data if there were three or fewer of these non-words checked. In other words 10% or less. If a tester checked four or more non-words we simply deleted the data. You always hate to believe data, so the other way that you can adjust is by using formulas, adjustment formulas. The problem is there is a lot of different formulas, and there is no real consensus on which one works best. In fact the different formulas have different characteristics that depend on the number of non-words that are checked, whether they are more severe or less severe. I think overall Yes, No tests work pretty well if few non-words are checked. But if a lot of non-words are checked, the estimates tend to be a bit ropery and we don't really have a good idea of the best type of adjustment formula to use.

Interviewer: That's really clear. So you've developed a new approach to scoring the Yes, No test that may overcome these problems. This involves reaction time. Can you explain what motivated you to look at reaction time and how this works in practise?

Respondent 1: Well we had this idea we were working on with all the reaction time studies, we had been working for a while on experiments using time lexical decision tasks, which is a [unintelligible 00:06:14] procedure in psycho-linguistics. For those who don't know what a lexical decision is, it simply consists of presentation of words, real words and invented words, one by one on a computer screen. And students have to say yes or no if they think the word presented is a real word. So it is slightly different from the Yes, No vocabulary test that Norbert was explaining before, so participants had to decide whether the word presented is the real word or not, and not whether they know the word or not, as is the case in the Yes, No test. As we were collecting data and working with the students we realised that when students knew words in a second, or foreign language very well they worked really fast in making their decisions and responding to the words. So this seems like a very logical assumption, the better you know something, so this could be a word, this could be an object, a picture, a landscape, someone's face, the faster you are at recognising it. But these seemingly obvious, and very logical assumptions hadn't been periodically investigated and this was our main motivation to conduct the studies report here in this paper.

It is true that in psycho-linguistic studies the relationship between accuracy and speed of responses has been widely explored, so there is a lot of research and studies in psycho-linguistics investigating the different factors that affect participant's responses in lexical decision tasks. We know things like the [unintelligible 00:07:50] the word is, the quicker your recognition. Or the more frequent the word is, the faster we recognise it, but as I said earlier, the level of vocabulary knowledge or how well you know the word hasn't been empirically related to the speed of response or reaction time. As I say, it seems plausible, and it is quite logical, but we didn't have evidence to prove it yet. So we decided to do it, we said okay let's take a vocabulary Yes, No test, instead of the more traditional lexical decision task, and let's see whether participants reaction time can tell us something about how well they know the words. If so whether we can use these kind of reaction time information to score the test, this is basically what we did.

Interviewer: We do of course encourage readers to look at your study in the journal for the technical details, which we really can't get into here, but you can summarise how you investigated the usefulness of response time, and what you found?

Respondent 1: The first thing we had to do was to check that our assumption, the assumption I explained in the previous question, was actually true. If we wanted to use reaction time information to score the Yes, No test, then we had to confirm first of all that there was a relationship between participant's reaction time and their vocabulary, not just the accuracy of their judgements. So we had to prove that the faster the response is, the more accurate the judgements were, whereas more [unintelligible 00:09:26] and responses would be slower, and this could lead to overestimation of their knowledge. This is what we did in the first study that we report in the paper. We wanted to know whether we could use this reaction time information to judge the accuracy of responses, and if so how we could do that. So we designed a computerised

version of the Yes, No test. Participants simply saw the words one by one in the middle of the computer screen, and they had to say whether they knew those words or not by pressing the yes no button in the box they had in front of them. And their responses were recorded, so we had two different types of information, whether they had claimed to know the word, and how long it took them to respond. So we had that information, and we now had to ascertain the veracity of participant's judgements. We had to examine whether participants were being actually honest when saying yes to a word, and whether reaction time information could help us in determining this veracity.

To do this we decided to conduct personal interviews with each of the participants. We knew that this would be time consuming but we wanted to be as accurate as possible in determining participant's actual knowledge of the words, and this is why we decided to do personal interviews. Then when we had the information from the interviews we compared participant's responses in the Yes, No test and the knowledge they had shown in the interviews, and what we found was that these connections between reaction time and accuracy of vocabulary knowledge did actually exist. When participant's claimed to know a word and they actually confirmed this knowledge in the interview, the response was significantly faster than when they were just overestimating their knowledge and they have said yes I know this word in the Yes, No test, and then they had shown no knowledge of the word in the interview. So we could see that there was a clear relationship between reaction time and how accurate participant's judgements were, so this was good news. But now what we needed to do was to calculate a point or a threshold in reaction time from which we could say okay these responses from these participants start to be very slow, and very probably they are over estimations of the knowledge, so we won't consider them as true responses, as true knowledge, even if participant's claim to know the word. And we used this to adjust test scores, so we tried different approaches to calculate this formula, and we finally came with the formula based on each participant's general reaction time in a standard deviation. And you have more detailed information about how we came to these formula in the paper. We had proved that there was a relationship between reaction time and accuracy of responses, and we had also calculated a way to do it, a formula that we could use to apply these reaction time approaches to the scoring of the test. What we finally did was to apply this formula, we recalculated the test scores, and results show that the reaction time adjusted the scores provided a very close match to the interview scores, which corresponded to participant's true vocabulary knowledge. So this is basically how we investigated the approach in very basic terms, but readers will find more detailed information in the paper.

Interviewer: And do reaction time score adjustments give more accurate results than the traditional non-word approaches?

Respondent 1: Well this is exactly what we did in the second study that we report in the paper. As we say there was no clear winner in the comparison. At the group level the reaction time approach did provide the best adjustment, so it worked really well. But then we also observed that at the more individual level there was more variation, and that the effectiveness and the advantage of one or the other approach depended on the number of non-words that had been selected. Also on the size of participant's overestimation, so how much participants were overestimating. So if fewer non-words were selected, the reaction time approach seemed to work really well. But then when more non-words were selected the other approaches seemed to work better. So this indicates that Yes, No tests might need to use more than one adjustment approach, but these will obviously take some ore research to see how this multi-adjustment approach works in practise.

Interviewer: To bring this podcast to a close, let's look at the impact of your research. Can you briefly say what you would advise Yes, No test designers to do in the future in the light of the findings that you have described for us?

Respondent 2: As Anne said, there was no clear winner in the adjustment approaches, and unfortunately from one study we don't really have any conclusive results that can tell people exactly what to do in all cases. But there are some things that came out. I think one of the most

important things that came out of the study is that if no non-words are checked it doesn't mean that test results are okay. I think that is really key. Now what do we do about adjusting, Anna mentioned multi-adjustment approach, that could be explored. We might do something like blend a Yes, No test with a certain percentage where the students have to demonstrate their knowledge of the items as some kind of check on the voracity. We also might give instructions that are a bit stricter in saying that be very, very careful in your judgement, although it is yet to be seen how well that actually works in making the students be more careful. And the inevitable call for more research. I think we need to look at checklists tests with other populations and try these different approaches, and different suggestions to see how they actually work.

Interviewer: Thank you very much for explaining your research for us. As we said last year there is clearly a lot more to do to improve the ways in which we assess vocabulary size. And your research clearly takes us further along the road. We are very grateful to you for taking the time to join us on Language Testing Bites.

Respondent 2: Thanks for having us, and we hope that our study stimulates more discussion about vocabulary assessment.

Respondent 1: And we really hope you enjoy the paper, and that you want to do more research in vocabulary testing. Thank you very much.

Interviewer: Thank you for listening to this issue of Language Testing Bites. Language Testing Bites is a production of the journal Language Testing from Sage publications. You can subscribe to Language Testing Bites 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]