

LTJ 28.4 Measurement Models

[Start of recorded material]

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

In Issue 28(4) of *Language Testing* we publish a paper by Mark Wilson, Professor in the Graduate School of Education at the University of California, Berkeley, on Measurement Models for *Language Testing*.

This is based on his Messick Memorial Lecture, which was delivered at the Language Testing Research Colloquium in Melbourne 2006. We invited Mark to join us on Language Testing Bytes to talk about his work and ideas.

Mark, I'd like to thank you for taking time out to talk to the readers of the Journal about your paper in Issue 28(4).

Respondent: Well, thanks, Glenn, it's a pleasure to be talking to you. It's was fun giving the lecture and it was nice to see it published, and this is an interesting follow-up.

Interviewer: Okay, well, many of the regular Journal readers will know what an item response model is but some listeners to the podcast may not. So, first of all, can you briefly tell us what an item response model is?

Respondent: Sure, it's a mathematical model but it relates to student ability to the probability that they will get a particular item right. Now, it's only an approximation to reality but what we want is a useful approximation of reality, and when it works it allows us to do some useful things. It's [UI 01:34] the core of it, the thing I think is most important, it allows us to map the student ability and the item difficulty, on to the same graph.

Interviewer: Now, in your article you make a distinction between descriptive and explanatory uses of item response models. Can you explain for us what the difference is between the two, and in particular define an explanatory use?

Respondent: Well, a descriptive model, and really in this context that's a traditional item response model, provides estimates for each item and each person of their location on that map that I just mentioned. Now, that's a very useful thing and we can do all sorts of, especially good teaching stuff with that. However, as we move on to explanatory models what they're trying to do is to explain why those locations are where they are using something that we know about either the person or the item. Some examples... an obvious example would apply very generally would be the grade of the student, the year that they're in in school. Equivalently an obvious one for an item might be the difficulty of the words that are inside the item.

Interviewer: Okay, can you give us some examples of the kinds of variables that might be used in an explanatory study, and tell us why the information might be useful to language testers?

Respondent: So, for students, the sort of things that you might find useful in language testing might be, for instance, the EFL status of the students, or something like the gender status, this will help us understand why particular groups of students are giving those particular responses.

We can look at it from the other side, from the item side, we can start thinking about what is it about the design of the item that has made it appear in a particular location. For instance, in language testing one of the things that might be important for reading, say, would be whether the reading, whether the text, was continuous or discontinuous, continuous of course being sort of typical prose, discontinuous being something like when you're reading some sort of tabular display like a TV guide or something of that nature.

Interviewer: Right, and in your article you provide readers with an example based on data from the International Schools Assessment for Reading, and you give an example of item properties as, reading aspect, text format and text type, and you write, and I quote here, “item properties are either manipulated within subject design factors, such as in our example, or they relate to an unplanned variation of the items, which is more common”, but isn’t it essential for any kind of explanatory analysis of the type you describe that variations are planned and documented in the test specifications?

Respondent: Well, I think what that statement reflects really is the state of the art, or the science of testing at this point. I think so far in testing we’ve been a bit like explorers. We’ve been sort of describing the geography of learning, you know, we find an item that seems to work, seems to be the thing we want to do, and we make copies of it and we cling on to it, learn how to change it a little bit. But I think really what we’re looking for are theories about the wording, sort of the tectonic plate theory of learning, and really these sorts of models are aimed at that sort of situation. So, my comment really related to the fact that at the moment a lot of tests they don’t have this sort of design built into them, so we can’t really expect to see a lot when we analyse them this way. I think we can get hints though and we’ve really got to spend some time I think working on the [UI 05:32], and then going beyond there and starting off from a design perspective.

Interviewer: The second half of the article explains three models, a person explanatory model, an item explanatory model, and a double explanatory model, can you briefly explain these for us?

Respondent: Sure. So, a personal explanatory model, this is one that’s aimed at explaining those person locations I talked about before, in terms of certain characteristics you might have of the person, and I mentioned a couple just a little while ago. On the other hand, an item explanatory model is going to do the same for the items, it’s going to try and explain the location of the item on the map using characteristics of the item. A double explanatory model does both at the same time, and the [UI 06:20] look inter-reactions between person characteristics and item characteristics, that’s really the interesting thing if we think about the double explanatory items, and then you get into topics like differential item functioning and so on.

Interviewer: Right. In your discussion of the item explanatory model you say that the variables of reading aspect test format and type didn’t well explain the variation in the item estimates. This reminds me of the long history of attempts to model item characteristics in this way, including rule space methodology and, more recently, fusion theory. All these have failed, doesn’t this really tell us that what we really need is better theories of test design before explanatory models can be useful, or are you more optimistic about what we might achieve using models like these in the future?

Respondent: Well, I certainly do agree that we need better theories of items, and of student understanding, and of how those can express through the items we extricate in the test. So, I often think we need... but it’s sort of the big theories of learning and we have a bunch of it, but what we don’t really have are the sort of theories that take those big theories and then put them in the context of particular items and so on, we don’t really have that micro-theory, and so I think that’s the sort of thing we need to work on.

And it really is something that’s got to have to do both with the learning theory and the characteristics and empirical and statistical features of items, so that’s why I think these models may be helpful. There have been other

models in the past, I agree, and I certainly see these models as taking, continuing that tradition. I'm hoping that these are more transparent than those other models. I don't want to go into the details of one model versus another because I don't think that's the point of the paper or this podcast, but I'm hoping that these are more transparent in the way they go from the standard models that we're using today, the IOT standard models, and building out from there.

Interviewer: And, is there anything else that, you know, you'd like to add that you think is kind of important for the readers to know, or things they might want to look out for, you know, things that are going to be coming up in the next few years, related to your research and what's going on in this area?

Respondent: Well, we certainly are trying to build these ideas into computer generated and computer scored items, and I think there's a lot of possibilities there. You still have to have those basic theories to start off with about the learning that's going to take place, and you still do have to have the sort of, as I said, these micro-theories about how the measurement actually works, but I think there are some possibilities of getting more out of it if we can be both delivering and scoring items in an automated way.

Interviewer: Can I ask one more, just one more question, and that is, do you think there is more room for, say, a content specialist and psychometricians to work together in this area, you know, people who are experts in this kind of modelling, together with the people who come from a language background?

Respondent: I think there's a great need for people from those two groups to work together, and I'm hoping that in the future we will certainly have people who are actually, whose expertise span both those areas, because I think it's, well, it's nice to be able to put together teams, I think it really takes the integration of the two to make the important advances that we need. And right now we're not really producing a lot of those guys.

Interviewer: Well, Mark, thanks for that last insight, and on that note I'm afraid we have to bring this issue to an end. Thank you very much for taking part in the podcast and for sharing your knowledge and research with us.

Respondent: Well, thanks Glenn, I enjoyed speaking, or having the opportunity to speak directly to the readers. It's an unusual approach, at least in the sort of areas I publish in, and I think it's a real good idea. Thanks.

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: [UI 00:00] Unintelligible