The State of the Art in LSP Testing:
An Interview with Dan Douglas

by Tim Newfields

Dan Douglas is a professor in the English Department's TESL/Applied Linguistics Program and Interdepartmental Program in Linguistics at Iowa State University. He received a MA in ESL from the University of Hawaii in 1972 and a Ph.D. in applied linguistics from Edinburgh University in 1977. He has taught at the over ten universities, including Hiroshima University and Chukyo University in Japan, and has served as an editorial advisor for Language Testing and the TOEFL 2000 Development team. This interview was conducted electronically in August and September 2001.

How did you first become interested in language testing?

I did my doctoral research in the area of reading ability among secondary school students in Botswana. To do the study, I needed to develop comparable reading tests in English and Setswana, the students' L1. I had the good fortune to have Alan Davies as my PhD supervisor, and he, being one of the leading language testing teachers in the world then and now, had a great influence on my development as a tester. I got very interested in a variety of cloze tests called clozentropy, which makes use of the concept of redundancy in language as a means to developing a scoring scale. I used this idea in my dissertation study and managed to publish an article about it in the Journal of Research in Reading in 1978. That initial "success" made me feel pretty positive about making a career in language testing.

Could you mention what clozentropy tests are in more detail?

Clozentropy is a scoring method, not a test format, so it can be used with fixed-ratio or rational deletion techniques - I suppose it could be used with C-tests, but I don't think it has. In any case, clozentropy is based on information theory and involves giving a cloze test to both a target population and a criterion population. The responses of the target group are weighted according to their compatibility with those of the criterion group - the more of the criterion group who give a certain response, the stronger the weighting of that response when given be a member of the target test group. The resulting total score is thus a measure of the linguistic compatibility of each member of the target group with the criterion group.

You mentioned the Journal of Research in Reading briefly. Probably most language teachers in Asia are familiar with only a narrow range of the journals which discuss language testing issues. What journals would you recommend for foreign language teachers who want to learn more about language testing?

Language Testing is the primary journal in our field. It comes out four times a year and carries articles on language testing research, new test development, and reviews of new books and tests. Subscription information is available at www.languagetestingjournal.com/. Incidentally, I'll become co-editor of that publication, along with John Read, in January 2002.

Another major source of information on language testing is the *Language Testing Update*, the official newsletter of the International Language Testing Association (ILTA). Readers should see the ILTA homepage for more information.

Could you briefly outline the main ways that your views on LSP testing differ from those of Prof. Davies?

Well, he argues that (1) we cannot make a theoretical case for the existence of language for specific purposes, (2) there are practical problems with implementing such tests, and (3) in any case LSP tests seem not to predict future performance much better than more general tests, but that nevertheless, specific purpose teaching and testing can be justified on pragmatic grounds that they appear to the test takers and score users to reflect the communicative tasks they are interested in performing in English. This, it seems to me, is really just a new version of the old "face validity" argument for the LSP enterprise - LSP is really just General Language dressed up in specific purpose window dressing.

I think this would be a mistake, and I argue that (1) language acquisition as a special case of a general capacity for language use, (2) both dialects and registers are learned and discarded as part of social behavior - this is related to the discourse domains hypothesis, by the way - and (3) are learned in contexts, so that the interaction between language knowledge and context changes the nature of both. Thus, it surely must be the case that there is such a thing as specific purpose language. I'm working on a paper to discuss these issues further.

One of the core concepts in your book Assessing Language for Specific Purposes *is the distinction between language knowledge and background knowledge. Who first introduced this concept? How widely accepted is it today?*

Well, you could start with F. C. Bartlett, who invented one version of schema theory in 1932. He was interested in how people use past experiences to organize new information, and in an experiment requiring subjects to reproduce geometric shapes they had seen only briefly, he found that they typically gave the shapes names based on their background knowledge - "two carpenter's squares," for example. Specifically in language testing, concern about background knowledge goes back to Lado in his ground-breaking 1961 book, in which he notes that professional or technical "meaning" that people need to do their jobs, "may not be used in tests," except, interestingly, in tests for members of a profession learning a foreign language - an early reference to LSP testing!

Could you briefly outline the "discourse domains hypothesis" which you developed with Larry Selinker?

We initially set out to explore variability in interlanguage production and the idea that IL performance varies with context. We soon developed the notion that language learners, as language users, create interlanguages within contexts of use, and that what counts in thinking about context isn't so much the external features (such as setting, participants, purpose, etc.) as the internal interpretation of them by the learners. This internalized context we called a discourse domain, and we believe that all learners - all language users for that matter - create many discourse domains to talk or write about aspects of their lives that are important to them or at least salient at the time, such as interacting with colleagues at work, schmoozing with friends outside of work, participating in classroom language learning activities, and taking language tests. . . until we understand how context, as discourse domain, works in language learning and use, we believe we won't fully understand second language acquisition and use. Interestingly, we called our first publication on discourse domains in 1985 "Wrestling with Context in Interlanguage Theory." Fifteen years later, in an update in the Annual Review of Applied Linguistics, Elaine Tarone called her article, "Still Wrestling with Context in Interlanguage Theory." It's a tough problem.

Can you mention why item response theory may often be unsuitable for LSP tests?

I think because, at least in theory, IRT procedures require that all test items or tasks be independent of each other, and often in LSP tests, we wish to develop complex tasks that are not independent, reflecting target specific purpose language use. In practice, though, IRT is probably robust enough that some task interdependence isn't a big problem, and the procedure is certainly useful for analyzing rater performance, for example, on production tasks. I don't think we've reached limits of what IRT can do for us yet.

You have mentioned how most LSP tests tend to have narrow assessment criteria with a strong linguistic focus. What advice would you give to those seeking to develop tests with a broader, more communicative assessment criteria?

It's a hard problem. We need to know much more about how people in the professional and vocational fields evaluate the communicative skills of colleagues and students in their fields. Such information will help us develop richer, more varied, and more relevant assessment criteria for our tests. As Tim McNamara puts it, we need to better understand "the nature of face-to-face communication as the basis for deciding on these criteria. In this we should be looking to colleagues outside our own field to help us understand..."

You mentioned a tendency of many institutions today to favor general tests over LSP tests. Why does this seem to be so?

Economics. It's more expensive to produce different tests for several disciplines than to produce one test that fits all. Also, there's been very little evidence that tests in different disciplines produce sufficiently different results to justify the cost. This has led testers and administrators to advocate more general purpose tests, for both theoretical and practical reasons. On the other hand, research has also suggested that the more discipline-specific tests are, the greater the variation in performance. So, if we really need to know whether a person can use a language well enough to work in a specific area, we need to develop a LSP test. Whether we need more or less specific tests always comes back to test purpose.

What changes do you see in the field of LSP testing? Can you mention why you feel optimistic about the field of LSP testing in general?

Well, to revive an old joke, a few years ago I didn't know what LSP testing was and now I are one! I think the biggest change is that LSP testing is emerging as a definable subfield of language testing at the same time as it's taking on its own identity as a branch of LSP. Specific purpose testing practitioners have had and will continue have a big influence on language testing in general, particularly our traditional emphasis on needs analysis as the basis for determining content and methodology. And, it has to be said, the language testing profession has had a beneficial influence on LSP testing by emphasizing the need for providing evidence for the validity of our tests rather than merely assuming that because a LSP test task looks authentic it must be valid. I'm optimistic about LSP testing because, as I look around at what's going on in the field, I see an enormous amount of vitality, creativity, and professionalism in every region of the world. I did an informal survey of current LSP test development projects for a "state of the art" paper and found interesting work going on in both academic and vocational LSP testing, from Finland to Guam, Hong Kong to Hungary, the US to Ukraine, and Japan to Saudi Arabia. LSP testing is alive and thriving.

HTML: http:// jlt.org/test/dou_new.htm / PDF: http:// jlt.org/test/PDF/Dougls.pdf