Voices in the Field:

An Interview with David Beglar

by Jeffrey Hubbell

David Beglar is the Director of the M.S.Ed./Ed.D. Program in TESOL at Temple University's Japan Campus. He received his Ed.D. from Temple University in 2000 and M.A. from Temple University in 1989. A resident of Japan since 1984, he serves on the JALT Journal Editorial Board, is a reviewer for *Language Testing*, and his articles in have appeared in *Language Testing, The Language Teacher, JALT Journal, TESL-EJ, and Reading in a Foreign Language*.   This interview was conducted via email in January 2009.

What first sparked your interest in language testing?

One of the first courses that I took as a Masters student in the late 1980s was a research methods course taught by Steven Ross in which we were doing t-tests and ANOVAs with hand calculators. I immediately liked the precision of the statistics and the fact that the statistics also let me know how much imprecision in each calculation was involved. This feeling continued as I learned more about the language testing field, as I appreciated how much statistics could inform me about item and person functioning. It helped take much of the guesswork out of test design. Basically, I think that my respect for the 'hard' sciences naturally led to me towards the quantitative side of the field of education.

Since you first got into language testing, what positive trends have you noticed?

First, noticeable progress has been in analytical techniques. For instance, I really appreciate Mike Linacre's (1987, 1989) development of the FACETS model, as well as structural equation modeling (Wright, 1921; Haavelmo, 1943; Simon, 1953). I think that both of models allow researchers and practitioners to understand test functioning more clearly. Second, I'm happy to see that awareness of the social aspects of testing has grown. Now most stakeholders are more keenly aware of how tests influence people's lives in important ways. Today we understand more clearly the need for critical selectivity about most kinds of assessment.

Conversely, are there any trends in language testing today that particularly concern you?

One concern is with the continued widespread use of raw scores as the basis of interpreting the results of high stakes tests and as the basis of research results. I think that Benjamin Wright (1989), and more recently Trevor Bond (2008), have made it clear that raw scores are potentially unstable and misleading and that they are best converted to interval measures using the Rasch model.

However, I think that on the whole, the professional language testing community is made up of many knowledgeable, honest people who are working hard to do what they believe is best for test-takers. As a result, I'm optimistic about the future of language assessment and feel that overall trends are positive.

What concept seems to be most difficult for novices to grasp about language testing?

Not surprisingly, the area of validity causes problems for anyone who really makes an effort to understand it in any serious way because it's so multi-faceted and complex. In the area of statistics, most people seem to have a great deal of difficulty understanding item and person fit statistics as well as dimensionality checks using item residuals. Personally, I think that understanding difficult concepts like these is a long-term project, so we all have to be patient and be satisfied with incremental gains over time. I'm still a student of the field of language testing in almost all ways.

If you had a magic wand and could change just 3 things about language testing in Japan, what would you want to change?

Everything I would change would be concerned with making testing and assessment fairer for students. First, I'd want all of schools using entrance examinations to train their examination committee in item and test analysis. Those tests need to be truly excellent because of their marked impact on students' lives. Second, I'd like anyone involved in high-stakes assessment to think about how to get more than one score from the test takers, since making high stakes decisions on the basis of a single score can easily lead to unfair results. The third change is related to the previous one: I'd like more test users to use the ideas of standard error and confidence intervals and to stop looking at test scores as if they're perfectly precise. We need to know which test takers are in the gray area between pass and fail and deal with them as fairly as possible.

For decades now, Temple University's graduate program has contributed significantly to the field of EFL and language assessment in Japan. In your opinion, what are its strengths?

One strength has been the outstanding people teaching assessment courses and seminars at Temple such as Thom Hudson, Steven Ross, J.D. Brown, Stephan Gaies, Lyle Bachman, Trevor Bond, and George Engelhard. And I'm very happy to say that Tim McNamara will be doing a seminar at Temple this fall. A second feature can be found in some of the works those persons have inspired. If you look at past issues of *Language Testing*, you'll find articles by a number of former Temple students such as Albert Dudley, Yoko Kozaki, Ed Schaefer, Gary Buck, and myself.

What do you feel are the most salient features of the Temple University TESOL program that other training programs might wish to emulate?

Two features stand out in my mind. First, we have always striven to find teachers who not only know their field well, but who also have a passion for teaching. We've always looked for individuals who are motivated to *teach*, not just to do research. Second, to the best of my knowledge, no TESOL program on earth has the number of experts teaching in its program that we do. That occurs because of the weekend seminars that we host nine or sometimes ten times a year. It's virtually a "who's who" of the field of SLA and includes names such as Rod Ellis, Mike Long, Paul Nation, Merrill Swain, Nick Ellis, and Gabi Kasper, to name just a few.

Do you envision any new trends for Temple's graduate program in Japan?

We are pushing many of our dissertation writers in three relatively new directions. First, we want people doing experimental studies to do multiple experiments, which has been the norm in cognitive psychology for years. Second, we are encouraging more mixed-methods studies in which quantitative and qualitative approaches are combined in a single study. Third, we are trying to get more longitudinal studies up and running, as development over time is an interesting and important issue in language acquisition. Finally, although Temple students have done well in the past, we are aiming to get even more of the dissertations published in major journals in the field. In short, our goal is to raise the general level of the research to even higher levels.

Works Cited

Beglar, D. (1995). Review of new ways in teaching vocabulary. *The Language Teacher 19* (2) 50-51.

Beglar, D. & Hunt, A. (1999). Revising and validating the 2000 word level and the university word level vocabulary tests. *Language Testing, 16* (2) 131-162.

Beglar, D. & Hunt, A. (2005, July). Six principles for teaching foreign language vocabulary: A commentary on Laufer, Meara, and Nation's Ten 'Best Ideas'. *The Language Teacher, 29* (7). 7-12. Retrieved January 9, 2009 from http://www.jalt-publications.org/archive/tlt/2005/ 07_2005TLT.pdf

Beglar, D. & Murray, N. (2002). *Contemporary Topics 3* (2nd edition). White Plains, NY: Pearson ESL.

Bond, T. (2008). If raw scores and Rasch estimates correlate so well, why bother? Value adding in test development and scoring with the Rasch model. Address at PROMS 2008 Conference. Aoyama Gakuin University: Tokyo, Japan.

Gorsuch, G. J. & Beglar, D. (2004, June). Teaching second language acquisition courses: Views from new faculty. *TESL-EJ, 8* (1) A-2. Retrieved January 9, 2009 from http://writing.berkeley.edu/TESL-EJ/ej29/a2.html

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11, 1-2.

Hunt, A. & Beglar, D. (1998, January). Current research and practice in teaching vocabulary. *The Language Teacher, 22* (1). 8-10. Retrieved January 9, 2009 from http://jalt-publications.org/tlt/files/98/jan/hunt.html

Hunt, A. & Beglar, D.(2005). A framework for developing EFL reading. *Reading in a Foreign Language, 17* (1) 23-59. Retrieved January 9, 2009 from http://nflrc.hawaii.edu/rfl/April2005/hunt/hunt.pdf

Linacre J. M. (1987). *An Extension of the Rasch Model to multi-faceted situation*. Chicago: University of Chicago, Department of Education.

Linacre, J. M. (1989). *Many-facet Rasch Measurement*. Chicago: MESA Press.

Simon, H. (1953), Causal ordering and identifiability. In W. C. Hood T. C. Koopmans, (Eds.) *Studies in Econometric Method*. New York: Wiley, pp. 49 - 74.

Wright, B. (1989). Raw Scores as Sufficient Statistics. Rasch Measurement Transactions, 3 (2) 62. Retrieved January 9, 2009 from http://www.rasch.org/rmt/rmt32e.htm

Wright, S. G. (1921). Correlation of causation. Journal of Agricultural Research, 20. 557-585. Retrieved January 9, 2009 from http://www.ssc.wisc.edu/soc/class/soc952/Wright/Wright_Correlation%20and%20Causation.pdf

HTML: http://jalt.org/test/beg_hub.htm / **PDF**: http:/jalt.org/test/PDF/Beglar.pdf