

Voices in the Field: An Interview with Glenn Fulcher

by Tim Newfields

Glenn Fulcher is a professor in the School of Education of the University of Leicester. He also co-edits the journal *Language Testing* and is on the editorial board of *Assessing Writing*. In 1993 he received his PhD in Applied Linguistics at the University of Lancaster and he also has a master's degree in Applied Linguistics from Birmingham. In 2003 he authored *Testing Second Language Speaking* (Longman/Pearson Education), and in 2007 he authored *Language Testing and Assessment: An Advanced Resource Book* with Fred Davidson. He has held various executive posts in the International Language Testing Association since 1998. This interview was conducted by email in July-August 2007.

It seems people get into the language testing field for various reasons. How you first become interested in this field?

Before I moved back to the UK I taught English in schools. When I was in Cyprus I was required to teach a lot of the examination classes. In fact, most of the levels in the school were linked to British tests. Talk about test-driven curricula! By far the most important was the old General Certificate of Education (GCE) English Syllabus 161b, produced by what was then the University of London Examination Board (later to become Edexcel). At that time this was the qualification for entry into British Universities. So most of my time was spent preparing students for that exam. And I had two problems with this. The first was how to actually teach my students what they would need to have successful academic careers in an English-medium environment, and develop the needed social communication skills when the test itself focused on the understanding of fairly literary passages and traditional 'composition writing'. The second was how to predict the grades that my students would get on the test. This latter requirement was pretty insidious. Teachers had to make predictions half way through the year; the school communicated those to parents and the examination board. I never did learn what the examination board did with the information, but parents certainly held teachers and the school to account. So you had to predict a grade that you thought a student would actually get, but not predict so low that it would demotivate the student or upset the parents so there was a complaint to the school, and not so high that there was a complaint after the results were published! In any case, my predictions were always well out, and I couldn't figure out why. But thankfully I didn't get too many complaints. Perhaps that was because my teaching wasn't entirely 'test oriented'. What I tried to do was look at the skills I thought they would need to do well in the UK, that were also required to do well on the test. I then developed my own materials and forgot about the test, apart from a periodic 'practice test' just for familiarization. This will probably all sound very familiar. Anyway, I started to doubt

whether my inability to predict scores was due to my lack of ability as a teacher, fluctuations in the abilities of my students, variations in test content, problems with the scoring method, and so on. As C. S. Peirce says, "the irritation of doubt causes a struggle to attain a state of belief. I shall term this struggle *inquiry*..." (1875, p. 114). And I was irritated enough to write up a research proposal and spend a number of years trying to find my answers. But of course, for every tentative answer there are many new questions. So this is what got me into language testing. Asking questions about the role of tests in my own teaching situation, how I related to the tests, and the problems that their use caused for me, my colleagues and my students.

From your perspective, how do you feel the language testing field is changing? What trends concern you most? Are there any you particularly laud?

It's always really instructive to go back and look at the periodic reviews of where language testing is, what it's achieved, and where we think it's going. These have been written by Davies, Skehan, Alderson, Bachman, and Alderson and Banerjee, among others. I also edited a special edition of *System* to look backwards and forwards, in 2000. What has really changed in the last few years is the perceptible emergence of a sense of confidence in the field. By this, I don't mean that we know the answers, that we're content with our methodologies, or even that we know what the key questions are. What I do mean is that there is more evidence that we're reflecting on our own practices, and the effects that our practices have. We're being more self-critical. And this is a sign of maturity. The critical language testing movement has been instrumental in this development, and Elana Shohamy's *The Power of Tests* comes across with almost as much force as Elana herself when she talks about test impact and our social responsibilities. There is real passion out there, directed at test misuse. One of the key targets is testing for immigration, and the use of language tests as the means to implement political agendas that would look rather ugly without the appropriate window dressing. And we now see the passion being tempered with a calmer consideration of the philosophical and ethical foundations that can underpin the reactions that we have to the way tests are used. The community is, I would like to feel, also becoming more tolerant because of this growing confidence. We can disagree, and learn from the disagreement. You probably wouldn't be surprised that I would at this point quote John Stuart Mill: "There must be discussion, to show how experience is to be interpreted. Wrong opinions and practice gradually yield to fact and argument: but facts and arguments, to produce any effect on the mind, must be brought before it." (1859: 25).

When we realize the impact that our work has on individuals all over the world it means that we have to address principles of test use, and also look back to question how the design of our tests is related to specific intended uses. One of the things that Fred Davidson and I have attempted to do is to show in our book and other papers, is that the social issues cannot be separated from principles of test design. And there is a great deal of work to be done in the nitty gritty of how we actually go about building tests well, so they are more likely to have the effect we intend and unintended effects can be identified and condemned.

This leads easily into what concerns me most. When large institutions start using tests or so-called testing frameworks to implement politically driven policies, I start to get worried. Collectivist solutions to perceived political problems have always resulted in oppression and the disenfranchisement of the individual; while this has

long been known in Europe (see Humbolt, 1854) we are now beginning to lose sight of the importance of protecting individuals in the undemocratic tendencies to impose harmony through standardization. The individual has no role in the new bureaucracy, other than as an efficient production unit in the economy of a would-be super-state fearful of its position in the pecking order of a global market. Tim McNamara and others are right to draw our attention to the ways in which tests and testers might be manipulated by policy makers, and to some extent using the framework of a social theorist like Foucault can help us see through to concealed intentions.

Although I certainly do not work within the tradition of Foucault, as you will probably gather from Unit A9 of our new book; and we explore this further in another new publication for 2007 in *Educational Philosophy and Theory*. Because Foucault's analytical method arises from a fundamentally pessimistic interpretation of history, it emphasizes the impotence of the marginalized and downtrodden to change their lot in life, and therefore any critique of testing from this perspective is of necessity negative. We work in a broader pragmatic tradition that critically analyses the intentions of the policy makers, the ethics of their enterprises, and the utility of the tools that they adopt or adapt. When the aims and intentions are undemocratic, or the solutions are essentially collectivist, there is a critical case to be made against them, as I have tried to do elsewhere, while recognizing that the tools themselves are potentially useful for other purposes. Unless such policies are challenged, what happens next is that no opposition is tolerated. As J.S. Mill also says, "All silencing of discussion is an assumption of infallibility" (1859: 22). And once collectivist solutions are in place they tend to spread and become more powerful, because they provide options for large-scale control of educational systems by unaccountable elites.

However, I'm essentially an optimist. I believe that a passionate concern for individuals, individual freedom, and the use of tests to provide meritocratic access to resources (only when absolutely needed) will prevail. And this is partly what professionalism and having good Codes of Ethics and Practice are really about. It's also why we need a strong International Language Testing Association (ILTA).

You mention some of the difficulties in getting a Code of Practice adopted by ILTA. What have been the main objections and what progress in that regard is being made?

The ILTA Code of Practice was finally approved at the LTRC meeting in Barcelona in June 2007. The Draft Code was subject to a number of amendments, perhaps foremost that it is now referred to as 'Guidelines' for Practice, but has nevertheless taken its place alongside the Code of Ethics. These documents will be available from the ILTA Web Site shortly.

The creation and adoption of the Code and Guidelines are due in no small measure to the tireless and expert work of Alan Davies. I'm sure your readers will be familiar with his extensive writing on ethics, codes and professionalism, one of which is reproduced in Unit B9 of our new book. He has an amazing grasp of the issues in developing professional codes and guidelines, from the philosophical to the legal. Of course, the road hasn't been easy. He and ILTA have had to take into account the variety in practice across cultures, which requires a degree of flexibility, without abandoning the core values that we as a profession wish to retain and promote. This

is a great achievement both for Alan Davies and for the language testing profession. It should be recognized as a milestone in the evolution of ILTA. Indeed, I also think it is the basis upon which affiliation to ILTA by regional language testing associations becomes completely meaningful for the world-wide language testing community.

Your most recent book emphasises the importance of effect-driven testing. Can you mention one or two examples of tests being developed in a way consistent with that approach?

We hope that this is an approach that will be used more in the future. However, to the extent that it provides a principled theoretical and philosophical basis for focusing our attention on test *purpose*, it is in fact consistent with that part of testing history that stresses the utility of tests for decision making within clearly defined parameters. The design of the test is therefore intimately connected with its intended use. We work backwards from a statement of use and purpose, to a design. This is where the metaphor of *test design as architecture* comes from, and therefore makes our work so compatible with Evidence Centred Design (ECD). It is just that ECD does not concern itself with anything beyond the immediate design considerations of the test, and so does not provide a larger philosophical and theoretical framework for test development and use. However, it is clear that the TOEFL® iBT, designed as it was using ECD principles on a clear statement of intended use: making decisions about readiness for study in North American Universities, so far comes the closest in my mind to an effect-driven testing project (Fulcher, 2005). It is no coincidence that we use a TOEFL Monograph from the iBT design research in Unit B6 to illustrate the process of linking prototype items to their intended effect. An interesting activity that someone might like to do, however, is look to see if they can identify tests that are likely not to have been designed on effect-driven principles. The greater the claim to generalizability of score use, the greater the lack of clarity and focus in score meaning, and therefore the larger the problem of investigating validity, as Micheline Chalhoub-Deville and I have recently argued (Chalhoub-Deville and Fulcher, 2003). Once you've identified a test and the claims made for score use, you can then ask if there is an associated research agenda to validate the score meanings for the stated test purpose. There are activities like this in our book, and on the website that accompanies it (see <http://www.routledge.com/textbooks/9780415339476/default.asp>).

Some readers might feel that effect-driven testing is perhaps too altruistic. In your most recent text, you acknowledge the fundamental profit motives behind the testing industry. If a test is making money, how can we count on test developers to try to limit the misuse of a given test?

That's a good question, and I guess that I can follow on from precisely where I left off. First of all though, I'm not entirely sure that I'm so cynical as to believe that making money is actually the fundamental motive in many cases. I do meet people who still believe with Edgeworth that testing "...is a species of sortition infinitely preferable to the ancient method of casting lots." And, we might add, "to the use of privilege or personal connections" (Edgeworth, 1880). Bernard Spolsky has done so much to make sure that we know about the history of testing (Spolsky, 1995), so we can understand how we have benefited from the past, as well as making us aware of the kinds of things we have to watch out for in the future. I'm particularly grateful for his rediscovery of Edgeworth because looking back on my own career, if it

hadn't been for the meritocracy that tests support, a kid from a working class background in the North of England like me would never have got to University. This idealism speaks to me, and I'm sure it does to many of my colleagues who are employed by testing agencies. It should be one of the motivations for doing this kind of work, and for striving to eliminate unfairness. But you're also right; this shouldn't blind us to the fact that some testing agencies have to make a profit.

The temptation to claim that your test can be used with any population, to test any construct, and make any decision, is very large. It's a growing problem in Europe too, where validity is often interpreted as a claimed link to the CEFR rather than relevance to clearly defined effects. It's similar to a belief that one drug will cure all illnesses. We know this is nonsense, but any one-solution-fits-all product is going to have higher sales for less investment. Until, of course, the customers realize that it isn't fit for purpose, and that there is no evidence that the scores are meaningful or can be depended upon for making fair decisions. What we need is a critical, open, questioning community of professionals, who are prepared to ask the difficult questions, and point out that the emperor has no clothes when it is necessary. Effect-driven testing, linking as it does use to design, provides a theoretical framework in which this can be done.

Your most recent book advocates an ongoing dialog among all test stakeholders. Ideally, each voice should contribute something to the ongoing debate. However, it seems the power differential between various stakeholder groups is quite wide. Large multinational testing corporations are almost a power unto themselves. How can mere test end-users actually gain a "voice"?

Your question actually raises some important issues. The first of which is whether the test providers are in fact a law unto themselves. There are organizations and governments around the world that have already decided what they wish to use tests for. A test is selected and then 'bent' to the purposes of the users, and if the provider does anything to upset the purposes and policies of the user, the test is dropped in favour of another provider who is compliant. There just hasn't been enough research into why some users 'select' tests; but actually getting them to talk about it would be very difficult indeed. On the other hand, there are institutional users who work with the test provider to define what it is they wish to use the test for, and a test is designed that does the task required, in an open and transparent way. It isn't all bad out there!

However, when you say 'end-user' I guess that you're not really thinking of an institution, although I'm afraid that institutions are most often the end-users. You're really thinking about the test-taker. These are the people who least often have any say whatsoever. But that is beginning to change, I think. Elana Shohamy has forcefully argued that test-takers should be consulted about the kinds of tests that they think are fair in their context. During test development projects the researchers are now asking test-takers how they reacted to the item types, whether the administration was stressful, and what could be done to improve the experience. Research into the impact of test-taker characteristics on scores is also important, as is the use of learners during beta-testing in prototyping. If you think back to test development practices even a few decades ago, there was almost no involvement of test takers apart from their use as 'subjects' to generate item statistics for analysis.

We've come a long way in a relatively short period of time, and the rights of the test takers are also protected throughout the research and test use process, by the ILTA Code of Ethics. In another ten or fifteen years we will have evolved further and found new ways to involve stakeholders, and the really good test providers will be part of solution to the problems and perceived inequities that we face today. Again, there are rogues out there who don't care about the end-users or test takers other than as customers, but these people are going to become the outcasts as we disseminate good practice both within and beyond the field. So without wanting to say that there aren't real problems, I'm still upbeat about what we as a community of language testers have achieved to date, and what we can achieve in the future. As long as we keep on talking about it - as we say in the book.

Since many of the readers of this publication are likely to be classroom teachers, what advice would you give to foreign language teachers with regard to testing?

That's a pretty broad question. But I guess that it brings me back to how I got into language testing in the first place. I'd have to say that teachers shouldn't let testing dominate what they do in the classroom. Above all, never get into the situation where all they do is practice for the test – any test – or use test materials endlessly. Good tests have manuals and other documents that describe what it is that the test is supposed to measure, and what the purpose of the test is. Teachers can use these and forget the actual test. They can get together with other creative people and ask: how do I structure the learning experiences and develop materials that will help my students to acquire the knowledge, skills or abilities that this test claims to measure? Then they can read the section of our book on reverse engineering, and start to look at items on the test: what do these test? What does a learner have to know to be able to answer this correctly? Then take this information and build classroom tasks that lead to the acquisition of these skills. Teachers should remember that test items and tasks are designed to elicit the maximum amount of information in the smallest amount of time possible, and that they are not connected to any other item or task in the test. What makes a good test item or task isn't necessarily what makes a good classroom task. This is what Units A2, B2 and C2 in our book are all about. My contention is that teachers who use their skills and creative talents will produce learners who can ace the tests, *ceteris paribus*, while those who are coached will do less well on the test, and after it! The same is true of classroom assessment. When I was an English teacher, I knew the capabilities of my students really well. Why? Because I'd worked with them for a year, sometimes for two or even three. I knew so much because they'd done linked tasks and activities that built on what had gone before, that were related directly to their current level of ability, and that were designed to push them one step forward. There was no division between teaching and assessment, in that teachers make ongoing judgments about learners in order to structure the learning environment. This is just a different paradigm from large-scale testing. Both have their place, but shouldn't be confused.

The problems really start when schools or educational systems introduce 'standards-based' approaches, because the purpose is no longer to assess for learning. It's about assessing for accountability, and this means that data have to be comparable across classes, teachers, schools, districts, or even countries. Once teachers are pushed into these accountability processes, much beloved by politicians and policy makers, classroom assessment is tainted and undermined. Teacher's assessments have to be

'aligned' to external standards, rather than the standards or frameworks being used as heuristics for teachers to think about their own assessment practices. So teachers also need to be aware of the political pressures on them, and resist them when necessary. Teachers might enjoy the scenario that is set up in Unit C8 of our new book, as it speaks directly to these problems.

The political mandate always moves for collectivist solutions rather than valuing the genuine achievement, progress and growth of the individual. Dewey was absolutely right when he first proclaimed individual growth as the main criterion for the judgment of educational achievement, and we need to remind ourselves that despite the onslaught of the collectivists this hasn't changed.

What plans do you have on the horizon?

Well, I have to continue developing the skills needed to get me through the next five years of co-editing *Language Testing* with Cathie Elder. I have to say that one of those skills isn't dealing with my co-editor! Cathie is just delightful to work with, as anyone who has worked with her will know. But cajoling colleagues to review papers when they're all worked off their feet is a different task. Proof reading really is a difficult skill, and I'm only just managing to slow down my reading enough to do it even tolerably well. Then there are the higher level decisions, either when reviewers disagree, which isn't often (thank goodness), or when there are questions of relevance or significance. Luckily, we have a great Editorial Board and two excellent immediate past-editors in Dan Douglas and John Read who are always ready to advise and share their wisdom.

As I've already indicated, I think that the work of ILTA is really important. I was President in 2006, and now I'm immediate past-President. The growth in ILTA membership and the expansion of the affiliation network is absolutely critical for the future of our profession, and that's something that I will be supporting.

There are a couple of test development projects that I'm involved in which will provide opportunities to use and refine some of the methodologies discussed in the 2007 book. And of course, my cooperation with my colleague and friend, Fred Davidson, will continue. Anyone who has read the acknowledgement in my 2003 book will know that Fred has had a major impact on my thinking, and his work on test specifications with Brian Lynch (Davidson and Lynch, 2002) is one of the modern classics in language testing. We met many years ago when we served on the same testing committee in the United States, and he got me into reading Peirce. As I already had a background in philosophy and have always enjoyed enlightenment writers, the empirical tradition, and the work of the 19th century liberals, J.S. Mill in particular, it was easy to see how American Pragmatism was part of this great intellectual tradition. And as you may know, William James dedicated his book *Pragmatism* to J.S. Mill. Many years later Fred visited me when I lived in Scotland, and we discovered that both the content of our bookshelves and our interests significantly overlapped. It was during that week in Broughty Ferry – the most beautiful and friendly town in the UK – that I asked Fred to join me in writing the book for Routledge. The collaboration and debate over the next few years were extremely creative and productive. We still have a number of papers that need writing up or revising and we plan to work on further projects together. We also plan that some of what we write will appear in the press as well as academic

journals, like the articles that we have published in the *Guardian*. It is important that issues in language testing that impact on people's lives are brought to the attention of a wider audience.

And finally, did you notice that the very last word in the text of our book is 'fun'? Perhaps it's the first time this particular lexical item has appeared in a book on language testing. It's important, both professionally and personally. Pragmatism implies a basically optimistic orientation to life. It holds that individuals do have the ability to make a difference, through their own work and their cooperation with colleagues. By working together and disagreeing together in an open, democratic community, we make progress as a community. So whatever else I do, I'll keep talking, disagreeing, challenging, and being optimistic that where we are now is better than where we were in the past, and that when I'm dead and buried (which I hope is a long, long, time away), things will be even better.

Works Cited

Chalhoub-Deville, M. And Fulcher, G. (2003). The oral proficiency interview and the ACTFL Guidelines: A research agenda. *Foreign Language Annals* 36, 4, 498 - 506.

Davidson, F. and Fulcher, G. (2006). Flexibility is proof of a good 'framework'. *Guardian Weekly*, 17th November. Available Online:
<http://education.guardian.co.uk/tefl/teaching/story/0,,1950500,00.html>.

Davidson, F. and Fulcher, G. (2007). The Common European Framework of Reference (CEFR) and the design of language tests: A Matter of Effect. *Language Teaching* 40, 3, 231 - 241.

Davidson, F. and Lynch, B. (2002). *Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications*. New Haven and London: Yale University Press.

Edgeworth, F. Y. (1888). The Statistics of Examinations. *Journal of the Royal Statistical Society*, 51, 599 - 635.

Fulcher, G. (2003). *Testing Second Language Speaking*. London: Longman.

Fulcher, G. (2004a). Deluded by Artifices? The Common European Framework and Harmonization. *Language Assessment Quarterly*, 1 (4) 253 - 266.

Fulcher, G. (2004b). Are Europe's tests being built on an unsafe framework? *Guardian Weekly*, 18th March. Available online:
<http://education.guardian.co.uk/tefl/story/0,,1170569,00.html>.

Fulcher, G. (2005). Better Communications Test will Silence Critics. *Guardian Weekly*, 18th November. Available Online:
<http://education.guardian.co.uk/tefl/story/0,,1645011,00.html>.

Fulcher, G. and Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. London: Longman.

Fulcher, G. and Davidson, F. (forthcoming, 2007). Tests in Life and Learning: A Deathly Dialogue. *Educational Philosophy and Theory*.

Humbolt, Von W. (1854). *The Limits of State Action*. Republished in 1993, Indianapolis: The Liberty Fund.

McNamara, T. & Roever, C. (2006). *Language Testing: The Social Dimension*. London: Blackwell.

Mill, J. S. (1859). *On Liberty*. Available Online: <http://209.10.134.179/130/2.html>.

Peirce, C. S. (1877, November). The Fixation of Belief. *Popular Science Monthly* 12, 1 - 15. Available Online: <http://www.peirce.org/writings/p107.html>

Shohamy, E. (2001). *The Power of Tests*. London: Longman.

Spolsky, B. (1995). *Measured Words*. Oxford: Oxford University Press.

JALT Testing & Evaluation SIG: 2007 Officers

Officers of the Testing & Evaluation SIG are elected for one year terms of office at the TEVAL SIG Annual General Meeting at each JALT Annual Conference.

Honorary Chair: J.D. Brown

Coordinator: Jeff Hubbell

Program Chair: Ed Schaeffer

Membership Chair: Gerry Lassche

Newsletter Editor: Tim Newfields

Website Editor: Tim Newfields

Treasurer: Mark Chapman



Last



Next



http://www.jalt.org/test/ful_new.htm (HTML)

<http://www.jalt.org/test/PDF/Fulcher.htm> (PDF)