

Dr Glenn Fulcher Talks to ELT

• Today's learners are considered 'consumers' who have the right to select the knowledge they wish to know. Is this shift a result of market needs and globalization? "The shift toward seeing learners as consumers has been going on for the last 15 years, but it has certainly accelerated in the last few years. Language is increasingly seen as a means to an end. This might be working in an international company where bilingualism is essential, entering an English medium university, getting access to scientific literature, or moving to live in another country. In all of these situations you have to show that you can use the language to work, study, learn or socially integrate. Over the same period the volume of language testing has increased exponentially, and new test providers have sprung up to offer both new and competing products. Learners have suddenly found themselves in a situation where schools have to sell courses that meet their particular communicative needs, and test providers need to market and sell the testing products the learners wish to obtain to achieve their goals. But there is nothing really new about this. Education and the certificates that come from taking tests have always been the gateway to a better future, and so they have always had a market value. You know, one of my favourite books on testing and assessment was written by Henry Latham in 1877, and I have a short quotation from page 6:

"People are hardly aware how thoroughly the educational world is governed by the ordinary economical rules."

Remember that this is the 19th century, and Latham argued that the reason students took tests was because of the economic value that was placed upon the outcome. He was talking about the kinds of tests that were taken at university and to get jobs in the civil service, and the test takers were just as much 'consumers' in the new globalized world of the British Empire as our students are today. What has really changed in our days is the huge expansion in testing volume that Latham couldn't have imagined, and the fact that it has become a global industry in itself. As you probably remember, at the end of 2009 Pearson launched its own tests to compete with Educational Testing Service and Cambridge ESOL, and the President of

*Dr Glenn Fulcher is Reader in Education (Applied Linguistics and Language Testing) in the School of Education, University of Leicester. He has published widely in language testing and assessment, and his latest book **Practical Language Testing** was published by Hodder Education in 2010. He has served as President of the International Language Testing Association (ILTA), and is currently the co-editor of the international journal *Language Testing* (Sage). You can discover more about language testing and assessment at his website: <http://language-testing.info>*

ELT NEWS met Dr Fulcher at the 32nd Language Testing Research Colloquium, which was held at Cambridge University last April, and did not miss the chance to talk to him about language testing.

Pearson Language Tests gave an interview in which he said: "It's a fairly commercial, competitive market already. We're going to make it more so." This was printed in the Global Business section of the New York Times (http://www.nytimes.com/2009/09/08/business/global/08pearson.html?_r=1), which gives you a clear indication of the motivation for getting into the large-scale testing business these days. Of course, this doesn't necessarily make what is happening in any sense 'bad' or 'irresponsible'. If there is a need for the tests in democratic societies that respect advancement by meritocracy, and if the development and delivery of these tests requires that we operate within markets, it may just be that the companies are giving the consumers what they want. The real issue is whether the testing products are well made and useful for the intended purposes."

• **Is testing a political activity in the sense that it is used to control educational systems?** "Testing has always been used to control educational systems. As I've said before, you don't have to look any further than Greece to see that this is the case. One of my favourite quotations is from *Plato's Republic*:

"we must see how they stand up to hard work and pain and competitive trials... And any Guardian who survives these continuous trials in childhood, youth, and manhood unscathed, shall be given authority in the state... Anyone who fails them we must reject."

Plato sets out the curriculum for the leaders very precisely, and the tests are there to ensure that the elite study what is required – the knowledge valued by the current elite, so that the social system is preserved. Of course, we have to remember that Plato was talking about a system that he would have liked to impose on Athens, but he never did. Luckily for us, Athens remained democratic and did not follow Plato's oligarchic advice. We see the tightest control over education through the use of tests in countries with extreme ideological systems and anti-democratic tendencies. The purpose is always the same – to make people comply with the ideologies of the elite, and to ensure that only those who do comply find themselves in positions of influence and leadership.

The classic example is the educational system imposed on Germany by the Nazis, which was totally out of keeping with the liberal educational and assessment practices that Germany had experienced for the previous hundred years. Nowadays, of course, we are lucky to live in much more benign political environments. But that doesn't mean to say that the tendency has gone away. We know from impact and wash-back studies that many governments use tests to try and change the way teachers teach, and how and what learners learn. The real reason for this today is the fear of politicians that we are in danger of losing our position in the global economic market. What they want are workers who have the skills necessary to be productive within our economy. I believe that this is one of the driving forces behind many of the cen-

tralizing 'top-down' initiatives in testing across Europe; politicians see the latest results of the PISA literacy tests and turn to what they see are 'simple' solutions. Of course, it's not that simple at all, but policy makers like 'simple'. I think that a similar motivation lies behind the recent attempt by the Council of Europe to make the Common European Framework the criterion for 'recognition' of qualifications across Europe. Teachers have a justifiably uneasy feeling that this is something that just isn't going to work, and I have written a great deal about why I think it can't and won't work. But when we listen to talk of 'standardizing' and 'harmonizing' our qualifications and educational system to produce a European educational zone that can compete in the world, we get a glimpse into minds that like the 'simple' solutions. Some of us think that diversity and choice are more likely to lead to economic productivity than producing the standardized cogs for European markets. Anyway, despite where we stand on current EU policy, the real point is that testing has always been used in policy making, because education has always - and probably always will - be seen by politicians as a means to change society in line with their vision of how we should live. This is stepping outside testing and into political philosophy of course, but my own view is that politicians should not attempt to imagine and work towards a utopia. They need to take a leaf out of Darwin's book – there will be all kinds of random and unexpected changes, some will be for the better, others won't. We hope that changes for the better will endure. But it is not



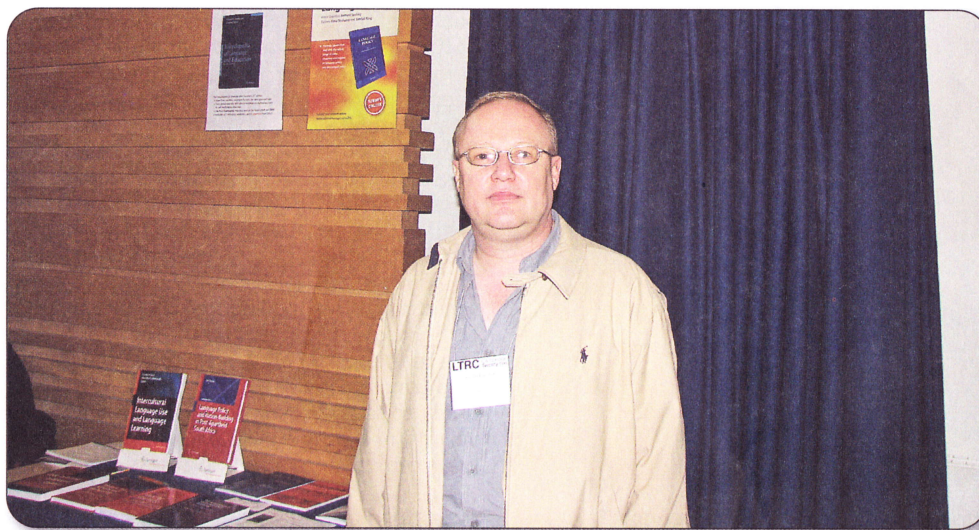
NEWS About Testing

something that we can control. I think that teachers are much better off being optimists than control freaks.”

• Are tests used to keep people out of countries and jobs?

“Yes, is the simple answer. Wherever we look in the world, countries are using language testing as part of their immigration policies. When a language test is designed for a specific purpose, the basic validation issue is whether the score from the test will support an inference to what the test taker can actually do or accomplish with the language. In turn, this inference supports a decision about his/her ability to perform successfully. Let’s take a high stakes example that is currently being talked about a lot. If someone wishes to get a job as an air-traffic controller we have to be certain that they can communicate with pilots and other controllers successfully in order to keep our airways, airports, and the travelling public, safe. Miscommunications account for many of the accidents and near-misses that we hear about in the news. Language tests are therefore devised that contain samples of the kinds of tasks that air traffic controllers do in real life, the types of language they need to use and understand, tested under performance conditions that they will encounter in the actual job. Of course, we can’t replicate everything, and that is why we have to make an inference from the score to what they may be able to do in real life. As part of designing the test, we have to make a cut score on the test, so that if someone gets higher than this score they can practice, and if they don’t, they can’t. What we don’t do is raise or lower this cut score once we have empirical evidence to suggest that it is the best cut score to make the decision.

Now, let’s compare this with the use of IELTS for immigration in Australia. In the last year the score required for immigration has been significantly increased in order to reduce the number of immigrants seeking trade employment. Australia just doesn’t need these immigrants during a global economic turndown. Most of them



come from India. But Australia doesn’t cap immigration. On the one hand, a blanket law regarding immigration would reduce immigration from everywhere, which they don’t want. On the other hand, targeted legislation would look like racism. But if you increase the English language requirements on the grounds that ‘higher scorers are more likely to integrate into our society’ the motivation can be presented as ensuring the wellbeing of all who come to our shores.

But when the economic crisis is over and additional workers are needed in restaurants, hotels, and other service industries, no doubt the language requirement will be reduced once more. Testing is being used a useful surrogate for an immigration policy, because it isn’t quite so easily classed as a xenophobic reaction to protect jobs, and it gives the authorities economic flexibility. From a testing perspective, we set cut scores based on validation evidence that tells us that people above or below a certain score can or can’t ‘do’ something with the language. We can’t just move the cut scores when we wish, without evidence. Now, if the Australian government actually commissioned research to show

that anyone below an IELTS 6 (for example) would have trouble integrating into society, this may (if you agree that integration is a ‘good thing’) be grounds for saying that someone can’t apply for a visa until they achieve that level of English. But this won’t be done, because it would mean they couldn’t lower the cut score again when they need more workers. Of course, Australia is just one example, and sometimes language testers use this example because of its history of using language tests for this purpose, dating back to the famous translation test of the early 20th century. But just look at how all European countries have put language testing in place to restrict movement and labour in the last few years. The whole issue really comes back to economics, which was where we were going in my response to your first question!”

• Is there such a thing as fairness in testing?

“My answer to this is unequivocally ‘yes’. Let’s go back to basics on this one. Tests of various kinds are there to provide evidence upon which decisions can be made. Think again about the example of the air traffic controller. We need to

make sound decisions because we want to be fair to the travelling public. We don’t want their lives put at risk. I don’t think anyone would argue with this. Let’s take another situation – getting a place at a top university. Not everyone can go to a top university, and not everyone can go to university. So how do we decide as a society who gets to go, and who gets to go where? Well, it all depends on what you value.

In a globalized market economy, perhaps someone would argue that what really matters in 21st century society is money. If we agreed on this, then we would let all universities charge what they wanted in fees. The very rich would get into the top universities, and the poor wouldn’t get a place at all. This is a statement of our values as a society. But this isn’t really acceptable to us. We know that there are going to be very clever poor kids, and there are going to be rich stupid kids. And in addition to that, we believe that people should be able to climb up the social and economic ladder rather than being trapped in some economic caste that they can’t escape from.

This was essentially the reason for introducing the British Civil Service Examinations in the 19th century. It was the dawning of meritocracy in modern Europe, although of course the Chinese had been using examinations for the same purpose for centuries. In short, being fair is about having a level playing field, where (*ceteris paribus*) everyone has an equal chance irrespective of how much money they’ve got, or which family they happen to have been born into.

Sam Messick, one of the greatest validity theorists of the 20th century, used the language of the legal writer Rawls to explain this as being related to ‘distributive justice’ – which is providing fair access to the limited goods that our world has to offer. So in testing, fairness is about everyone having an equal chance, ensuring that the test isn’t biased against anyone, making sure that someone who is disabled in some way doesn’t get a lower score just because of the disability, and that someone who is prepared to cheat doesn’t get away with it. The reason I can’t imagine a world without tests is because I can’t imagine any other social instrument to provide this kind of fairness. So whatever we think about tests and testing, we just haven’t thought of any other solution to the fairness issue.”

• A lot of teachers teach to the test. Is this healthy? †12

Dr Glenn Fulcher Talks to ELT NEWS About Testing

♦11

"I guess it all depends on what you mean by 'teaching to the test'. We all know that one of the goals of many learners is to pass a test. As teachers, we have to ask why they want to pass the test. The test itself is just gateway to something else – most often getting into university. If the test is any good (by that I mean that there is validation evidence to show that the scores are useful for the intended purpose) then what it tests should reflect the language and communication skills necessary for university study. If 'teaching to the test' means 'practising test items', or 'answering past papers', then it is extremely unhealthy. The only purpose to doing past papers is to familiarise learners with the test format so that the item types aren't a problem for them when they do the test. Once they're familiar with the test, teachers should be teaching the language, skills and abilities necessary to function well at university and, incidentally, to pass the test. If they can do the latter, the test is no longer a problem. I would go even further than this. Any teaching practice called 'teaching to the test' which succeeds in raising test scores to pass levels, while having no effect on the learners' language abilities, is unethical. It undermines the purpose of the test, and it condemns the learners (should they pass) to a life of misery once they arrive at the university and find that they can't write their assignments or communicate with their peers or teachers. On the other hand, teaching which genuinely results in improved language learning and communicative ability, which is reflected in higher test scores, is not a problem at all."

• Has technology improved the way languages are taught and learnt?

"Well, I have to confess that you're talking to a technophile, so I'm going to say that it has. Clearly, the most important technological development is the internet. It provides materials and opportunities to listen and communicate that we couldn't have dreamed off even ten years ago. Classes can take place in virtual learning environments, and the advent of Second Life and similar virtual environments is making



this even more attractive for younger learners. We all know that there are dangers too. But used sensibly by creative teachers, these new technologies have a great deal to offer.

When it comes to testing and assessment, computer based testing is some way behind what we can do in teaching and learning. But there is a reason for this, and it goes back to the idea that testing is about providing a level playing field when making decisions. Whenever we are unsure whether a new medium might give some sub-group an unfair advantage we don't start to use it until sufficient research has been done. These fairness issues lead to caution. This isn't necessary with teachers in classrooms, where they can try out new technologies without worrying about fairness. In fact, it is because teachers are more adventurous that new technologies are slowly incorporated into assessment.

For example, back in the 1990s there was a lot of research done to find out if asking learners to type essays in tests would disadvantage those who were not familiar with using computers. The assumption was that using a pen and paper was 'normal', and using a keyboard would be unfamiliar. If you remember, the first

computer based TOEFL test had a tutorial teaching test takers how to use a mouse and keyboard before the test started. That's all gone now. The assumption is that computer familiarity is widespread, and that perhaps asking people to write with a pen is going to cause a disadvantage.

At the moment there is a lot of research being done on the use of video in tests, as opposed to audio files, asking whether watching helps or hinders comprehension for some people. We'll get to a stage where this is no longer an issue as well, and then we'll move on to something else. There is nothing to fear from technology. It's all useful in the right hands. The real questions haven't changed – is the medium useful to all, and if it isn't, are we being unfair to some? These are questions about the validity of the meaning of scores, which is always central to testing and assessment."

• Are computer-based tests as reliable as pen and paper ones?

"A very good question, and I'm assuming that you are referring to writing and speaking, rather than multiple choice type tests, because on these the computer is simply more efficient. So when it comes to a computer scoring writing, the answer is 'yes'. But, there is a 'but'. What do we mean by 'reliable'? I'm assuming that you and all your readers will use word processing software on a regular basis. Probably you use Microsoft Word. When you've finished typing something and you review it, you can get a word count. For the same piece of work it will always give you the same word count. Similarly, if you get it to check your grammar and spelling over and over again without changing anything, you'll get the same comments. That is perfect reliability, as the definition of 'reliability' is 'consistency'. Compare this with a human counting the number of words, or giving feedback on 'errors'. Humans aren't good at counting words, and you'll get two counts from two different humans. One person will also mark things as errors that another won't, and so on. However, is "isn't" two words or one? Your software says it's one.

A teacher would say it's two. How about this: "too many cooks spoil the broth". Is that five words, or is it a single lexical chunk? There was a great story in the UK press last year when the examination boards were considering using a computer to rate advanced level English essays. They gave the computer some of Churchill's speeches: "We shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the fields and in the streets." The computer gave it a low mark, saying it was "repetitive and below average." So there's a much more important question, which is: "how do we know that the computer is giving a score that reflects the language skills and abilities of the writer?" This brings us back to the issue of validation and score meaning again. All of the companies that produce computer scored tests of writing correlate the computer scores with those assigned by human raters. These correlations are generally speaking fairly high, often achieving rates of agreement of between .7 to .8. The evidence shows that the computerers agree with the human raters as well as how separate human raters agree with each other. But this does not mean they are sensitive to the same features of the text. Computers count words, they can analyse the vocabulary used, they can look at cohesion, and so on. What they cannot do is make a judgment about the communicative quality of the writing, or make an inference about the appropriateness of the genre for a particular audience. In short, computers are great at counting stuff, and if the scores correlate with humans, I don't have a problem with an essay in a test being scored by a human and a computer, rather than two humans. What I do have a problem with is an essay being scored by a computer alone. On the evidence we currently have, and our knowledge of what the computer is capable of doing, I cannot see any justification for allowing a computer to make a decision about language use, which is going to change a person's life. It's back to fairness and validity once again."

Anastasia Spyropoulou
(anastasia@eltnews.gr)