Is the TOEFL exam aimed at everyone? Research considerations in the training and application of the TOEFL exam abroad

Jesús García Laborda
Universidad Politécnica de Valencia, Spain

Jesus.garcialaborda@uah.es

Abstract

In recent years, the Educational Testing System organisation has developed two models of the computer-based Test of English as a Foreign Language (TOEFL). However, the computerization of the test has shown a number of problems according to the testees' origin. This paper suggests some of these problems after conducting short interviews with four TOEFL teachers; i.e. problems in delivery, interface design and test taking contextual validity (Weir, 2005). These aspects and those presented in other educational systems need to be addressed in further research. The fact that TOEFL is only a compulsory test for those who wish to pursue further education in the USA cannot prevent ETS from ignoring the potential problems of the current model in different parts of the world. Thus, further research on this topic needs to be carried out in Europe and elsewhere. This paper also serves as a starting point for countries and institutions that may be considering the implementation of computer or Internet-based applications for high-stakes testing such as university entrance examinations.

Keywords: CALL, Internet, testing, university entrance examinations.

1. Introduction

This paper intends to suggest the importance of addressing the cultural differences and national educational policies in relation to standardized tests, especially in TOEFL. According to García Laborda (2007), many high stakes tests have become computerized in the past decade, and some others are also planning to follow this direction. As a consequence, students need to address their study of second language (or foreign language) learning in different ways since test strategies and exam–examinee relationships have changed dramatically. The credibility and international recognition of the TOEFL exam largely depends on a fair application of rigorous measures of construct and content validity. To ensure the correct performance of the TOEFL test in the United States and abroad, ETS devotes a considerable amount of its budget to research that (as opposed to IELTS) can be accessed free of charge on the company's website. This research, which is generally well planned and based on large samples, usually fails to be statistical and tends to look at the exam processes from a receptive perspective. A receptive perspective would be that of the incoming larger culture or linguistic location. For instance, when students take the TOEFL exam they need to adapt their language and testing skills to a North American format. This means that they may perhaps need to change their learning skills and their approach to test-taking and adapt to what the test administrators believe is most appropriate in terms of testing knowledge. Of course, a high stakes exam like this has its own philosophy on how students will have to use the language at the time of the test and once they are admitted into an English speaking academic environment. However, this perspective tends to ignore the emitting countries' conventions, constraints and uses of foreign language teaching. Since the construct of TOEFL is very clear, well defined (and refined) and is in a constant revision process, it would be difficult to deny that the test matches its administrators' purposes. However, is it not a fact that students need to adapt themselves to the test rather than being assessed or diagnosed fairly in a global educational world?

Opposed to a utilitarian approach, which has been mainly supported by those who establish a difference between assessing English as a second language and English as a foreign language (Al-Musawi, 2001), there is another trend of researchers and teachers who believe that students need to be assessed of their language skills despite the use that the foreign language will have. However, since this paper intends to address the TOEFL test, these two opposed perspectives will be left aside. Nevertheless, the first ethnocentric approach certainly affects the candidates who register for the test every year. Furthermore, this point of view affects two main matters that have become the landmarks of the exam since 1998 such as computer literacy or perception of audiovisual repertoires that are included in the test. These matters obviously affect students according to their language and culture. This paper will suggest ways in which the current TOEFL test may discourage Spanish students from taking the exam as a part of the requirements to pursue university graduate and undergraduate studies in the United States. This research points out two main aspects that may affect TOEFL; on the one hand, the possibility of some students choosing to take other diagnostic tests such as IELTS or any other that may fit the students' cognitive style more appropriately (see a list in García Laborda, 2007); and on the other that, although TOEFL is a non-profit organisation, it encompasses a large number of linguists, office assistants, researchers and workers that may be affected by a lower number of candidates than expected.

2. Why is TOEFL harder for the Spanish candidates?

2.1 The TOEFL format

Since 1998, the TOEFL test has progressively undergone a number of changes. That year the Educational Testing Service (ETS) began to computerise the traditional pen and paper test. Of course, a number of trials and studies were conducted to observe whether it was possible to automate the test and if the results were similar.

At that time reports showed that when students were computer literate the differences were minimal, and consequently the Computer Based TOEFL (CB TOEFL) (Netten, 2000; Banerjee & Clapham, 2003) was implemented and operative by 2001. Progressively, the test items became richer and an audiovisual repertoire was included as prompts in the test. The items supported by audio recordings soon included videos as well (Ginther, 2002).

In the first years of the new millennium ETS started to consider going online, and finally in September 2005 the new Internet Based TOEFL became operative. Apparently, the new TOEFL (as it is commonly called) had a number of constraints at the beginning but these constraints have apparently been solved. Today, the iBT TOEFL coexists with the traditional pen and paper test although it is expected that sooner or later the pen and paper TOEFL will be discontinued.

The new TOEFL includes productive reading, writing, listening and speaking items and, more importantly, a combination of them. Today, TOEFL is the only standard high-stakes test to include all four skills. The new TOEFL comprises the following sections:

TOEFL iBT	Description
Listening/Reading Sections	Listening passages mimic college lectures and follow more "natural" speech patterns, including pauses, corrections, and grammatical errors. Reading passages are more academic in nature to better reflect college reading assignments, and students have access to a glossary.
Structure Section	Grammar skills are tested throughout the test.
Writing Section	There is one 30-minute, and one 20-minute essay.
Speaking Section	Six open-ended speaking questions require test takers to speak into a microphone.
Integrated Language Skills	Some sections of the iBT combine four basic communication skills. For example, a test taker might listen to a lecture and read a passage, then write or speak about it.

Scoring	1-120 (human graders for essay and speaking section).
	Online score reporting allows students to view their scores at the end of the test.
Time	3.5 hours

2.2 How are Spanish students affected by the iBT TOEFL format?

As mentioned in the introduction of this paper, ETS takes careful measures to ensure the validity of the test. In fact, a large number of researchers devote part of their research to analysing the test and its candidates form different perspectives. Unfortunately, most of them are quantitative and based on large and, more often than not, only on statistical analysis that supposedly supports ETS's methods and procedure but provides little information to the general practitioners (especially because most of these reports and papers require an advanced knowledge in statistics). As a consequence, except for large groups of students, generally from Asia, cross cultural studies are far more frequent than qualitative studies centred on just one nationality, and how differently examinees from different parts of the world behave in the test.

The key issue of this paper is how Spanish students are affected by the iBT TOEFL format. At this point, I am unable to know the exact number of Spanish students who take the TOEFL exam every year but certainly although not the largest, it is not the smallest either. One of the issues that very few testing organisations have researched is the state of education in the candidate's country of origin. In the case of Spanish students, there are relevant factors that may certainly bias or modify their results. The Spanish educational system requires between 9 and 12 years of study of a second language, mostly English. Additionally, students entering the university are required to take a test of the same foreign language. However, because of the way English is taught in primary and secondary education and the washback effect that this has on the way English is taught during the last two years at school is the

reason why there are no specific oral requirements and written production is limited to short compositions on general topics with which the students are acquainted. Another issue that may account for inferences in this issue is the students' level of computer literacy. Although some studies have addressed this issue, university students in Spain do not seem to show the same computer capabilities as other students in Europe or in the United States.

As a result of these factors that are found (and maybe even blamed) on the educational system, students taking the TOEFL exam may experience the following constraints:

- Constraints in equating their results across the four skills,
- Constraints in adapting to computer environments,
- Constraints in adaptation to test interfaces,
- Constraints with the classroom setting.

Constraints in equating their results across the four skills: One of the major concerns for ETS is finding ways in which the results can be representative of the L2 knowledge and then equated to obtain just one mark for the test as a whole. This has led ETS to move from adaptive testing (in CB TOEFL) to regular testing in which all the items are considered evenly (iB TOEFL). Likewise, Spanish students who according to the PISA report are among the lowest achievers in foreign languages and verbal expression in Europe may face difficulties especially when writing and speaking while they may excel in grammar (which is unfortunately greatly emphasised in Spain). Thus, many Spanish students may show great discrepancies between the different language cognitive skills. This fact may not only have serious consequences on Spanish students but also on students from other origins (Takanashi, 2004).

Constraints in adapting to computer environments: The second issue is whether students all over the world have similar computer

skills. Previous reports by Manalo & Wolfe (2000) and Taylor, Eignor & Jamieson (2000) have suggested that computer skills vary according to different academic environments and cultures. Reports in Spain show that students have limited computer skills, and although they may prepare their papers using word processors (mostly MSWord or Open Word), they also rely on programme-provided support that they are not likely to have in a test situation (for instance, spelling checkers or grammar tutors). Additionally, students taking the iBT TOEFL are required to use an American keyboard and although there seem to be no studies focussing specifically on this problem, probably adapting to the new keyboard means slowing typing speed and maybe having to monitor one's typing more often and undertaking different writing strategies. All of these aspects may certainly affect the student's performance negatively. Moreover, some students may be more familiar with computer use in academic settings than others because the degree of computer implementation varies significantly between schools, local and regional boards of schools, etc. TOEFL administrators should also pay attention to the fact that students who feel more confident working with computers will certainly develop more and stronger learning strategies adapted to this environment, and will benefit from using computers in their tests more than the other candidates.

There is also another issue relating to computer use that is currently scarcely addressed; i.e. anxiety towards the use of computers. Since some students' performance with computers may be below average, they may feel threatened by the presence of other class mates in the training courses and thus intimidated by having to take the exam in a context in which they "feel alien". So far, ETS has done a good job by allowing students to choose between the computerised or the pen-and-paper versions of the TOEFL test but if the intention in the future is to eliminate the latter progressively, will all the students be tested only on their language knowledge and skills?

Constraints in adaptation to test interfaces: García Laborda & Magal Royo (2007) suggest that interfaces can establish differences between users and cultures since the perception of interfaces is clearly influenced by the natural reading movement (right to left or left to right), visual culture (eastern versus western) or even cognitive styles. Additionally, colours and fonts may also influence the student's adaptation to different interfaces but these are difficult to account for except for students with certain limitations (such as colour-blind candidates). Besides, the reading layout should also be considered. Lonsdale, Dyson & Reynolds (2006) consider that the way in which the reading elements are presented may produce significant changes in the score of a reading test.

Constraints with the classroom setting: For Muroyama (2006), students tend to adapt their ways of learning to the format in which they will be assessed at the end of a period of instruction. Teachers also tend to adapt their teaching practices to that format (washback effect; Alderson y Hamp-Lyons, 1996; Green, 2005), as well as ETS introducing significant changes in the classrooms where students are to take the tests as an institutional policy. These changes tend to benefit the students but they establish high demands from the instructors and, in fact, TOEFL teachers now do not only need to be language experts but also be acquainted with the use of ICT in the classroom, otherwise secondary problems that the students may not be able to handle (due to their own limitations in computer use in academic environments) may arise.

Conclusions

Students of different cultures may face different sorts of constraints when facing the pen-and-paper and iB TOEFL exam (Leeson, 2006). ETS has made limited efforts to account for these difficulties and has kept a rather conservative attitude expecting students to change rather than making an effort to understand the candidates' problems and peculiarities. In fact, statistical research plays a significant role in justifying this conservatism in their educational policies. The

idea that if foreign students intend to study in the United States they have to be tested according to what they are going to see and live in the academic and everyday life in the US seems old fashioned and to a certain extent authoritative, especially when students are supposed to prove their English knowledge and not their overall language, academic and computer skills (otherwise, why have tests like the Graduate Record Examination?).

Although the iB TOEFL exam is well planned and developed it may perhaps not be efficient in assessing a number of students worldwide who may have to choose another test in order to be accepted into North American higher education institutions or who may have to take internal tests organised by larger universities to facilitate their graduate students transition into the North American academic life. Thus, the co-existence of the pen-and-paper (and oral interview) version along with the new iB TOEFL test still seems necessary. Maybe shortly students from most countries will be able to take online tests but, up until now, this may not prove to be possible. If ETS does not want to miss educational, research and commercial opportunities, they certainly need to consider their examinees' realities and approach the problem of origin with an open mind rather than just a justifying thought of all the goods that these new iB TOEFL will certainly bring. In this sense, this paper is only a first approach and a call for more cultural and anthropological research that will, no doubt, benefit institutions and students alike.

References

Al-Musawi, N. (2001) The validity of scores on TOEFL and FCE for predicting students' success at the university. *Dirasat: Educational Sciences*, 28(1), 117-126.

Alderson, J. C. y Hamp-Lyons, L. (1996) TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.

Banerjee, J., & Clapham, C. (2003) Test review: The TOEFL CBT (computer-based test). *Language Testing*, 20(1), 111-123.

García Laborda, J. (2007) On the net: Introducing standardized ESL/EFL exams, *Language learning and technology*, 11(2), 3-9.

García Laborda, J. y Magal Royo, M. T. (2007) "Diseño y validación de la plataforma PLEVALEX como respuesta a los retos de diseño de exámenes de idiomas para fines específicos", *Ibérica*, No. 14, Fall 2007, 79-98.

Ginther, A. (2002) Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, 19(2), 133-167.

Green, T. (2005) Review of washback in language testing: Research contexts and methods. *Language Testing*, 22(4), 539-545.

Lee, Y., Breland, H., & Muraki, E. (2005) Comparability of TOEFL CBT writing prompts for different native language groups. *International Journal of Testing*, 5(2), 131-158.

Leeson, H. V. (2006) The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, 6(1), 1-24.

Lonsdale, M. d. S., Dyson, M. C., & Reynolds, L. (2006) Reading in examination-type situations: The effects of text layout on performance. *Journal of Research in Reading*, 29(4), 433-453.

Manalo, J. R. & Wolfe, E. W. (2000) *A comparison of word-processed and handwritten essays written for the test of English as a foreign language*. East Lansing: ERIC document n° ED443845.

Murayama, K. (2006) "Adaptation to the test": A review of problems and perspectives. *Japanese Journal of Educational Psychology*, 54(2), 265-279.

Netten, G. (2000) TOEFL introduces computerized exam in Asia. *American Language Review*, 4(5), 40-41.

Taylor, C., Eignor, D., & Jamieson, J. (2000) Trends in computer use among international students. *TESOL Quarterly*, 34(3), 575-585.

Jesús García Laborda
Universidad Politécnica de Valencia, Spain