Web-based tests in Second/Foreign Language Self-assessment

Miguel Fernández Álvarez
Juan García Rico
Cicero Public Schools

Abstract

This article focuses on the use of computers in Language Testing. The first part of the article gives a general introduction to three different kinds of tests administered with a computer: (1) computer-based tests (CBTs), (2) computer-adaptive tests (CATs), and finally (3) web-based tests (WBTs). On the other hand, the second part of the article focuses on self-assessment and web-based tests, presenting the DIALANG project (a new European work which has developed diagnostic language assessment tools in fourteen European languages, delivered over the Internet).

Introduction

Technology is increasingly being an important factor of change in education. As a matter of fact, instructors, administrators, students, parents and book editors and writers have experienced during the last two decades the greatest revolution that ever happened in the classroom. This advancement has been influenced basically by the introduction and use of the computer in the process of teaching and learning. Thus, learning became significantly richer as students started having access to new and different types of information, as they manipulated it on the computer through graphic displays or controlled experiments in ways never before possible, and as they could be able to communicate their results and conclusions in a variety of media to their teacher, students in the next classroom, or students around the world.

The rationale behind Foreign and Second Language instruction is based on the notion that maximum contact with the target language yields maximum learning, which is why this field has found a great advantage in the introduction of computers in the classroom (Council of Europe 1997). The new approach allowed addressing all four critical skills in language learning (speaking, listening, writing, and reading) by engaging the students' interest and providing a culturally and linguistically realistic context, at the same time that more materials were accessible.

Computer-based tests (CBTs)

Language Testers also found a new perspective for the assessment and evaluation of language acquisition. In fact, CBTs started being developed and used since the early 80s. These tests represented a new way of measurement for language competency, and were delivered on an individual computer or a closed network. CBTs can be offered at any time unlike mass paper-and-pencil administrations which are constrained by logistical considerations. In addition, CBTs consisting of dichotomously-scored items can provide feedback on the test results immediately upon completion of the test. They can also provide immediate feedback on each test taker's responses, a characteristic that is very useful for pedagogical purposes (Bennett 1999). Here we highlight some of the main advantages found in CBTs:

- Efficient Administration: CBTs can be administered to individuals or small groups of students in classrooms or computer labs, eliminating timing issues caused by the need to administer paper/pencil tests in large groups in single sittings. Different students can take different tests simultaneously in the same room.
- Immediate Results: One of the major drawbacks of testing on paper has been the long wait for results because of the need to distribute, collect, and then scan test booklets/answer forms and hand score open-response items and essays. On the other hand, the results of computer-based tests can be available immediately, providing schools with diagnostic tools to use for improved instruction, and states with information to guide policy. Even open-ended items can be scored automatically, greatly reducing cost and scoring time (Thompson 1999).
- Efficient Item Development: As computer-based testing becomes more developed, item development will be more efficient, higher quality, and less expensive (National Governors Association 2002). Bennett
(1998) believes that at some point items might be generated electronically, with items matched to particular specifications at the moment of administration.

- Increased Authenticity: Computers allow for increased use of “authentic assessments”—responses can be open-ended rather than just relying on multiple choice.

**Computer-adaptive tests (CATs)**

Adaptive testing is not new. In fact it has probably been around for centuries as a better way to test. And even since the beginning of the 20th century, when large-scale testing began, adaptive tests were some of the first tests ever constructed. Computers as a testing medium rapidly attracted the attention of psychometricians because they allow the application of Item Response Theory (IRT) for delivering CATs, which can often pinpoint a test taker's ability level faster and with greater precision than paper-and-pencil tests. The computer can make the necessary calculations needed to estimate a person’s proficiency and to choose the questions to present. Based on the test taker's responses, the computer selects items of appropriate difficulty thereby avoiding delivering items that are too difficult or too easy for a test taker, but instead selects more items at the test taker's level of ability than a non-adaptive test could include (Alderson, Clapham & Wall 1995; McNamara 1996; Fernández Álvarez and Sanz Sainz 2004). CATs “offers a potentially more efficient way of collecting information on people’s ability” (Hughes 2003: 23). Figure 1 shows in a clear and detailed manner the way CATs work:

![Selection of items in a CAT](image)

Figure 1. Selection of items in a CAT

According to Davies *et al.* (1999: 29), CATs “are claimed to facilitate greater accuracy of measurement as a result of presenting candidates with items which are at the maximum discretion level, i.e., are more or less at the candidate’ level of ability, items of this type providing more information about the candidate than items which are too easy or too hard.”

CATs, like CBTs, show some advantages. In fact, CATs were developed to eliminate the time-consuming and traditional test that presents easy questions to high-ability persons and excessively difficult questions to low-ability testees. Dunkel (1999) identifies some other advantages in their use:

- **Self-Pacing:** CATs allow test takers to work at their own pace. The speed of examinee responses could be used as additional information in assessing proficiency, if desired and warranted.
- **Challenge:** Test takers are challenged by test items at an appropriate level; they are not discouraged or annoyed by items that are far above or below their ability level.
- **Immediate Feedback:** The test can be scored immediately, providing instantaneous feedback for the examinees.
• Improved Test Security: The computer contains the entire item pool, rather than merely those specific items that will make up the examinee's test. As a result, it is more difficult to artificially boost one's scores by merely learning a few items or even types of items (Wainer 1990). However, in order to achieve improved security, the item pool must be sufficiently large to ensure that test items do not reappear with a frequency sufficient to allow examinees to memorize them.

• Multimedia Presentation: Tests can include text, graphics, photographs, and even full-motion video clips, although multimedia CAT development is still in its infancy.

Web-based tests (WBTs)

The last innovative testing technique is the so-called WBTs, which is a CBT delivered via the internet, written in the "language" of the internet (HTML), and possibly enhanced by scripts. The test is located as a website on the tester's server where it can be accessed by the test taker's computer, the client. The client's browser software (Netscape Navigator, MS Internet Explorer) displays the test, the test taker completes it and -if so desired- sends his/her answers back to the server, from which the tester can download and score them. If the test consists of dichotomous items (true/false or multiple choice) it can be made self-scoring by using scripts.

Even though we can still find several drawbacks in the use of WBTs (like security or technical problems such as server failure and browser incompatibilities), it is obvious that there are several advantages linked to them. Roever (2001) distinguishes three main advantages:

• Flexibility in time and space: This is probably the biggest advantage of a WBT. All that is required to take a WBT is a computer with a Web browser and an Internet connection (or the test on disk). Test takers can take the WBT whenever and wherever it is convenient, and test designers can share their test with colleagues all over the world and receive feedback.

Even though security is a big issue in the use of WBTs, there are still advantages to delivering the test via the Web, that is, no specialized software necessary, existing facilities like computer labs can be used as testing centers.

• Easy to write: Whereas producing traditional CBTs requires a high degree of programming expertise and the use of specially-designed and non-portable delivery platforms, WBTs are comparatively easy to write and require only a free, standard browser for their display. In fact, anybody with a computer and an introductory HTML handbook can write a WBT without too much effort, and anybody with a computer and a browser can take the test; language testers do not have to be computer programmers to write a WBT. This is largely due to HTML's not being a true programming language but only a set of formatting commands, which instruct the client's Web browser how to display content. In addition, HTML contains elements that support the construction of common item types, such as radio buttons for multiple-choice items, input boxes for short response items, and text areas for extended response items (essays or dictations).

• Affordability: A WBT is very inexpensive for all parties concerned. Testers can write the test by hand or with a free editor program without incurring any production costs except the time it takes to write the test. Once a test is written, it can be uploaded to a server provided by the tester's institution or to one of many commercial servers that offer several megabytes of free web space. Since WBTs tend to be small files of no more than a few kilobytes, space on a free server is usually more than sufficient for a test. The use of images, sound, or video can enlarge the test considerably, however, and may require the simultaneous use of several servers or the purchase of more space.

CATs are possible on the Web and do not pose many technical problems beyond those encountered in linear tests but it cannot be emphasized enough that the design of a sophisticated WBT is a very complex undertaking that requires considerable expertise in IRT.

The DIALANG Project

In this section we will focus on the use of WBTs in Second/Foreign Language Self-assessment, based mainly on the DIALANG project. (www.dialang.org).

DIALANG is one of so many examples of WBTs. It is a new European project which has developed diagnostic language assessment tools in fourteen European languages, delivered over the Internet. The fourteen languages of the DIALANG project include all the official EU languages as well as Irish, Icelandic and Norwegian. The assessment materials cover all levels from beginners to advanced; there are separate assessments for reading.
writing, and listening as well as for structures and vocabulary. The project complements traditional approaches to language assessment by exploring the use of new technologies for assessing the use of language.

DIALANG offers validated tests of different language skills, together with a range of feedback and expert advice on how to improve the skills. It also offers scientifically validated self-assessment activities and allows users to determine their language level, strengths and weaknesses as well as to increase their awareness of current skills and of what it means to know a language. It has been developed by more than twenty major European institutions, with the backing of the European Commission and is based on the Council of Europe's Common European Framework (CEF) of Reference (2001), which has become established throughout Europe as the most widely recognized frame of reference in the field of language learning.

The purpose of the DIALANG assessment system is to provide language users and learners with diagnostic information about their language proficiency which will help them to become aware of their strengths and weaknesses and to find ways of improving their proficiency (Alderson 2005). The assessment system will also help learners locate themselves in relation to the levels of the CEF, for a variety of uses specified in the Framework and associated documents. Feedback from the system will also provide learners with information that can be used as a guide when setting new learning goals and planning their future language learning.

Like other WBTs, it offers many advantages such as the ones mentioned above: flexibility in time and space, easiness to write and affordability. As opposed to proficiency tests, the main advantage that it offers is its security: it would be useless to memorize lists of vocabulary, or to take the test several times in order to learn the specific items, because the basis of the test is self-assessment. The test does not provide a certificate to enter a university, or to obtain a better professional position. Trying to cheat the machine would result into cheating oneself.

Furthermore, DIALANG offers the test takers the possibility to choose the skill, skill level, sub-skill, topics, etc. that they wish to assess. All this learner oriented control means that the test is designed for a large variety of learners. Any candidates preparing for any kind of examinations, following any kind of teaching methodology, and aging 16 and over could benefit from the test assessment.

The main gain of the project is probably the way in which it provides assessment information:

- The feedback is immediate.
- Scores are objective.
- There is not a global grade but an identification of the test taker’s current level.
- It suggests learners’ further study.
- It enables the planning of curricula.
- It tries to cover learner’s needs
- It allows the storage of the results for later comparison to check progress.

The procedure to follow in order to take the test does not require a lot of computer skills. The first thing to do is to choose the language in which instructions will be given. Instructions can be chosen in a different language from the language tested in order for the student to understand them whatever his/her level of proficiency in the target language is. After that, the skill and language to be tested are chosen. Then, test takers have the option to take a placement activity to measure their vocabulary knowledge in the target language and determine the items that should be present in the assessment. The next step is an optional battery of self-assessment questions that the learner has to answer in order to receive some feedback about what he/she thinks his/her level is. Next, learners take the test on the skill chosen previously. The items are frequently multiple choice questions, but sometimes a word or phrase has to be written. In these cases, only one answer is possible. Finally, DIALANG displays the results in terms of level, answer verification, score in the placement test, self-assessment feedback and advice.
The window under “your level” shows the level of the learner’s proficiency according to the CEF. Under “advice”, information about the current level is provided such as the capabilities of the learner, and recommendations for further development that would facilitate the step into the next level are granted.

In figure 3 it can be seen the information that DIALANG provides about the test takers’ level of proficiency indicating the type of speech that they understand, what they generally understand and their conditions and limitations.
A real challenge for CBTs, CATs and WBTs is to assess productive skills: speaking and writing. DIALANG lacks at the moment a system to assess listening and writing in terms of full sentences, paragraphs or essays. However, a project to include these deficiencies is currently taking place (Lancaster University n.d.). These are some of the items that are in the experimental phase:

- Indirect speaking with audio clips as alternatives. After listening to a recording, the test taker has four options which are also recordings intended to be the answer or reply to the first recording. This technique simulates speaking since the answer is not written, but oral. At the same time it is a multiple choice item, as we can see in figure 4.

![Figure 4. Indirect Speaking with Audio clips as Alternatives](image)

- Interactive picture with sound. Oral instructions are given and the learner has to give an answer interacting with the computer. In figure 5, instructions are given to locate a room in a map. The answer is not the traditional written multiple choice item, but locating the room in the map.

![Figure 5. Interactive Picture with Sound](image)

- Re-organization. In this activity some sentences have to be rearranged in order to make a coherent sentence (figure 6). Test takers have to drag and drop the sentences until they find the order they consider correct. This technique is taking us to a closer step towards the evaluation of coherence and cohesion through a computer.
Benchmarking in direct writing is the opportunity to assess free writing. Learners are given a prompt, and they have to write freely about it. After they finish their text, they are given a range of sample texts to compare with their own. Figure 7 provides an example. The sample texts are answers that could be given by a learner in a determined CEF level. As a final assessment tool, advice is provided stating the features that characterize each kind of writing level and some tips to help improve writing skills.

Self-assessment in speaking. A question is posed for which several oral answers are given. The test taker has to decide whether he/she can do it better than the recordings. Based on the answer, a level is assigned to the reply and feedback information about the level is provided (figure 8).
Conclusion

The importance of assessment in Second/Foreign Language learning is unquestionable. The feedback received from assessment can be used to place a student in the correct classroom, to measure learners’ progress, to revise and develop curricula, to create material, to propose teaching methods, etc. That is why it is not surprising that so much attention is focused on improving the assessment system. Technology plays a very important role in the development of testing methods. There are different kinds of computerized assessments such as CBTs, CATs and WBTs that offer great advantages to assessment research: They provide immediate feedback, ways to store test results for further analysis, improved security, storage of large amounts of items, grading objectivity, multimedia presentations, test takers’ self pacing, etc.

However, there are certain areas that pose a real challenge to computer testing. Computers lack human intelligence to assess direct speaking ability or free written compositions. It is here where research is studying new pathways to improve assessment. DIALANG is a current project that aims these goals. DIALANG researchers are designing innovative test items to diagnose learners’ levels of proficiency, and the improvements are having a considerable impact on the considerations of computer testing.

DIALANG focuses on self-assessment taking into account the learners’ interest in grading him/herself fairly, but it could not be used as a certification system. No wonder that future research will take into account the progress achieved so far and will try to apply it to other kind of tests, such as placement tests, certification systems, etc. Furthermore, although we cannot expect computers to equal human intelligence, DIALANG is demonstrating that there are ways to improve computer testing. We will be expectant to hear new advances that will surprise us as much as DIALANG latest findings.

References


