Applications

Adaptive Item Language Assessment based on students' cognitive abilities

Hara Giouroglou and Anastasios A. Economides

University of Macedonia

hara@uom.gr

The promotion of multilingualism, communication, mobility and cross-cultural awareness among EU member-states has created the demand for easily-administered, self-paced, time-effective, multilingual, and internationally accredited foreign language assessments (FLA). The Common European Framework of Reference (CEF) has paved the way by setting internationally certified standards for formal testing as well as self-assessment in all European languages and describing in detail the productive and receptive skills needed to attain a specific level of competence. CEF is based on the communicative, action-oriented and skill-based approach to language learning which is the essence of linguistic competency (Council of Europe, 2001, 2004).
The new generation of assessments acknowledges the fact that there is no average student model with predetermined behavior and tests are adapted to students' diverse educational and socio-economic background, age, nationality, first language, motivation and temporal accessibility. In order to foster learners' success, we need to adapt FLA environments to accommodate learners' diversity accordingly. Any assessment in foreign language that does not adapt to the aforementioned mixed student abilities cannot be considered reliable and valid. Mixed abilities create mixed needs which result in mixed implementations in all educational settings. This paper will describe the development of AILA, a computer adaptive and adaptable placement test in English as a Foreign Language (EFL) for mixed-ability students that can measure productive and receptive foreign language awareness efficiently.

Edizioni Erickson – Trento
www.erickson.it

Je-LKS Journal of e-Learning and Knowledge Society — *Vol. 2, n. 3, novembre 2006 (pp. 367-372)*

1. Introduction

There is a perceived need for a new generation of FL tests, which should be adaptive and adaptable in nature, catering for diverse, mixed-ability students. FLA needs to adapt to individual student needs, abilities, backgrounds, strengths and weaknesses, giving emphasis to cognitive language skills, such as comprehension, production and use. Computer Adaptive Language Testing (CALT) should also incorporate the CEF standards in order to develop internationally accredited, valid and reliable assessments. The Adaptive Item Language Assessment (AILA) is an adaptive placement test, based on CEF standards, that is both adaptive and adaptable in that examinees are given the choice to select how to answer each item presented. The system adopts course content tailored to the student's needs, taking into account different difficulty levels as well as different knowledge levels. AILA will be shortly ready for pilot testing.

2. Adaptive Item Language Assessment (AILA)

2.1 The Problem

Traditional FLA implementations fail to cater for mixed-ability students, as they are linear and targeted to the average student. CALT technology can provide student-centered assessment, replacing traditional testing wherever possible. However, CALT is based on solid programming which is collective rather than individualized and fails to include crucial cognitive parameters of student language competence and performance. Such systems cannot replace the human examiner without detrimental consequences for its group of examinees. The new generation of assessment systems for cross-cultural examinees should not assess students horizontally as an equable lot but vertically as mixed-ability individuals with mixed-scoring options. Moreover, the new generation of assessments can motivate test-takers, as it is proven that the new technologies are preferred by students (Ali, 2001).

The majority of CALT systems use multiple-choice (MC), close-ended items to distinguish proficient, good and weak learners. This is mainly due to the fact that MC items are easily programmed and calibrated in Item Response Theory (IRT). The program can easily identify correct and wrong answers and move on to easier or more difficult items. This technique is also reliable and valid as long as items are adequately pre-tested and correctly calibrated, even though the validity of multiple-choice testing has been seriously criticized (Chapelle, 2001). However, MC items cannot allow active expression and language production. Examinees are passive viewers of the proposed answers and they only try to distinguish the correct answer from the distracters. This method is widely used by language testing organizations, such as the University of Michigan Certificates in English,

while other organizations use a variety of MC and open-closed items, such as the Cambridge Syndicate and the State Examinations on Language Competence. Proficient learners answering MC items are not given the opportunity to separate themselves from good learners by openly typing the correct answer. They have to choose from the four intended choices and receive the same mark as other learners who may have accidentally chosen the correct item. This limitation does not allow the proficient learner to be discernable from others by testifying active language production. Another problem is caused by the prohibition of item reviewing. In psycholinguistics there is a clear distinction between errors, made due to ignorance, and mistakes, made due to negligence. Examinees are prone to mistakes not only out of ignorance but also out of misunderstanding, anxiety, confusion, distraction or other physical reasons. Since item reviewing is impossible in CALT, adaptive systems may form false impressions and give low scores. To this end, CALT should become more «intelligent» and simulate the human examiner in order to produce more accurate and precise scores.

2.2 System Description and Adaptation

AILA measures the ability of non native speakers of English to use and understand English as a foreign language for achievement and placement purposes. The test-takers who sit AILA can quickly assess their competence in English in the scale issued by the Common European Framework of Reference (CEF). The test measures competence in four out of the six CEF levels: A2 - pre-intermediate, B1 - intermediate, B2 - upper-intermediate, and C1 - advanced. Each level has an even number of Grammar, Vocabulary and Reading items.

We used a CPU Pentium III 800 MHz, with 2 GB RAM, and the Apache web server. The software can run on Windows NT 4.0, Windows 2000 Server or Advanced Server and the system software includes the required MySQL database software. For reasons of re-usability XML has been used to separate content from the way it is processed (i.e. presented) and avoids re-writing the same content that needs to be displayed in different formats. The software used is Windows 2000, My SQL (free), PHP, VB script, Javascript, HTML, XML.

AILA is computer-based, using adaptive technology in item selection. The system increases student motivation, by providing tailored content adapted to individual needs and level of competence. This maximizes the students' gain whilst reducing the time equivalent. The User Profile keeps standard information of each test-taker and is also updated each time the test is administered.

AILA is also adaptable in that the student defines his/her level of competence according to which the test will be administered. Secondly, as the test is administered, examinees are given the freedom to choose how to answer each item. They can either type an open answer, by typing their answer in the gap (OA) or

choose the correct answer in MC mode. The proficient examinee can type the answer in the OA, demonstrating his/her advanced knowledge. A correct OA response receives a bonus in the total score (standard grade + 0.25) and updates the User Profile of the examinee. Then, the item selection algorithm proceeds to the next item of increased difficulty. A wrong OA response immediately directs the examinee to the MC mode of the same item and the score is dependent on the MC scale. When the MC mode appears, the examinee cannot go back to the OA mode. Wrong choices in the MC mode receive no mark and the next item is easier. This method does not affect the final score of the test or punish a wrong OA, but it promotes examinees to demonstrate productive FL use and active FL extraction from their long-term memory. The duality of the system is adapted to students' divergent cognitive strengths and weaknesses.

AILA also tries to distinguish between errors and mistakes, using a simple method. It is a fact that in most MC questions at least one destructor is so close in meaning or in grammatical resemblance to the correct answer that may puzzle even examiners. The MC destructor that bears a close resemblance to the correct option is regarded as «incorrect but acceptable», receives no mark on the total score and the next item is of equal difficulty, giving the examinee a second chance to demonstrate competency on the same level. In the OA mode the system can also understand common speech errors, known as «slips-of-the-key», such as anagrams (e.g. «nad» instead of «and»). These types of mistakes are very common both in first and foreign language production, especially under stressful conditions and time limitations. Bearing in mind the fact that language is a flexible, ever-changing, living entity used to communicate meaning and retrieve information, we should not severely punish answers that have a slight deviation from the standard form.

2.3 Item Bank and Stopping Rule

The item bank consists of 600 items divided in the four CEF levels of competence (A2, B1, B2, C1), signifying item difficulty (b1-4). In each broad level of competence, items are sub-divided in three discriminatory levels (a1-3): The first discriminatory level (a1) contains items that are expected to be answered correctly by all examinees having the given competence, the second (a2) contains items that can be answered correctly by the average examinee, while the items in the third level (a3) can only be answered by the most competent students in this level. Finally, each discriminatory level is seperated in 5 content areas (c1-5), in order to ensure that examinees will answer a wide variety of language items.

The test starts with a given difficulty specified by the test-taker (bx), low discrimination (a1), first content area (c1), and random item selection. If the test-taker answers in MC mode correctly, then the next item is of the same difficulty (bx), medium discrimination (a2), second content area (c2), and random item

selection, otherwise the next item is one difficulty level lower. If the examinee answers in OC mode correctly, then the next item is of higher difficulty (bx+1), high discrimination (a3), fourth content area (c4), and random item selection. In this stratified way, we ensure that examinees will gradually attain their level of competence by answering different item types. AILA algorithm has a compulsory minimum number of 15 required items. Thus, the minimum test length is 15 items and the maximum is 40 items. The test stops when the examinee answers at least 15 items, having shown competence at one level of difficulty. There are no time limits per item; however, the maximum test time is 45 minutes.

3. Outcomes and Conclusion

FLA for mixed-ability students should be personalized, flexible, and sensitive to human cognition, language processing and error correction. AILA, an adaptive placement test which measures competence in EFL in terms of CEF levels, gives students the chance to show productive and receptive language use. The system also tries to discern errors from mistakes by evaluating students' answers. Thus, proficient learners will be able to excel, showing active language production. AILA has simple technological features that ensure ease of use and navigational transparency. The system provides on-line help in the examinees' native language on how to process every item during test administration, thereby minimizing possible confusion for the test takers. The test items examine basic language skills which focus on syntax, grammar, semantics and sociolinguistics, and use authentic language whenever possible. The test can be easily administered and updated by examiners, who can delete or add items whenever needed. An elaborate statistical itemisation and student analysis system collects information regarding items' administration and test-takers' performance. In future, AILA aims to more efficiently exploit the potential of the computer by adding multi-media elements to the application.

BIBLIOGRAPHY

Akmajian, A. et al. (1998), Linguistics. An Introduction to Language and Communication. Third Edition. The MIT Press.

Ali, A. (2001), «Technology Integration and Classroom Dynamics» In Proceedings of AACE Webnet 2001 World Conference on the WWW and the Internet, October 23-27, Orlando, Florida, USA, pp. 7-8.

Armstrong, T. (1999), Seven Kinds of Smart : Identifying and Developing Your Multiple Intelligences. New American Library.

Chapelle, C.A. (2001), Computer Applications in Second Language Acquisition. Foundations for Teaching, Testing and Research. Cambridge: CPU.

Chomsky, N. (1997), Powers and Prospects. Reflections on Human Language and the Social Order. Second Edition. Pluto Press.

Council of Europe, (2001), Common European Framework of Reference for Languages. Cambridge University Press.

Council of Europe, (2004), Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment. December 2004. Language Policy Division, Strasbourg.

Gardner, H. (1993), Frames of Mind: The Theory of Multiple Intelligences. Basic Books.