

大學英文網路評量與學習檔案系統之建構

A Web-Based Assessment and Profiling System for College English

Zhao-Ming Gao 高照明

zmgao@ntu.edu.tw

Department of Foreign Languages
and Literatures 台大外文系
National Taiwan University

Chao-Lin Liu 劉昭麟

chaolin@nccu.edu.tw

Department of Computer Science
National Chengchi University
政大資科系

摘要

我們發展一套針對大學英文閱讀, 聽力, 寫作, 與字彙能力的半自動網路評量與學習檔案系統。此系統除了能當作網路測驗的工具, 也能從質與量兩方面分析學生英文的優缺點與問題所在。

關鍵字: 網路評量, 網路測驗, 學習檔案系統, 語言評量

Abstract

We develop a web-based assessment and student profiling system which can evaluate and analyze students' academic performance in English classes at university level. The system is designed to qualitatively as well as quantitatively assess different language skills such as vocabulary, reading, listening, and writing. Apart from being used as a tool for semi-automatically creating web-based tests, it can also help keep track of students' shortcomings, errors, and progress.

Keywords: web-based assessment, web-based testing, student profiling, language assessment

Introduction

With the advent of the Internet technology, web-based testing (WBT) systems have played important roles in eLearning. Unlike traditional computer-based testing systems, web-based testing systems have the advantages of geographical, temporal and platform independence (cf. McCormack and Jones (1998), pp. 19-22). The system reported in this paper has more features than a web-based testing system. As with all CBTs, it has the capability of generating instant score reports, immediate feedbacks, and detailed item analysis. In addition, it has functions which most CBTs lack.

By far, item banks in most computer-based language testing (CBLT) systems are manually created by human experts. This procedure is not only laborious but also makes CBLT system inflexible, because once an item bank is manually created, it is difficult to adapt to the individual needs of each tester. To maximize automation and flexibility, item banks should ideally be created through semi-automatic means. Taking flexibility and automation into account, our system can semi-automatically create item bank for vocabulary tests and tests for reading speed. Moreover, it encompasses evaluation based on students' writing. By decomposing language skills into fine-grained traits using

qualitative and quantitative methods of assessment, the system can identify students' difficulties, weakness, strength, and progress.

The Pedagogical Aims and Architecture of the System

The system has two pedagogical aims: (1) to facilitate effective and efficient assessment (2) to create beneficial washbacks. For these two purposes, several web-based tools are designed to help teachers easily create tests and keep track of their students' learning activities. It is hoped that the profiling system can make students aware of their own achievements and weakness and hence can create beneficial washbacks for their learning.

The system architecture is designed with three rationales in mind, automation, modularity, and adaptability. The design philosophy of the system is that it should offer testers maximum convenience by reducing their tasks to simple operations. It should also be flexible enough to allow further refinements without the need to change the whole system. Modules should be designed to be reusable in different types of tests.

The system consists of four independent yet interrelated modules: (1) the item generation module (2) the test delivery module (3) the record keeping module (4) the data analysis module. The item generation module supports both manual and semi-automatic creation of items. The test delivery module allows teachers to choose the class to be tested, the type of test, the difficulty level of items, the number of items, and the time limit of each item to be answered. It also supports random generation of items from item bank. The record keeping module records all the items,

students' responses to each item, students' submitted homework, and students' learning activities that are deemed important. The data analysis module computes item difficulty, analyzes students' errors, and shows the weakness and strength of each student.

Assessing Reading

Three aspects of reading ability are assessed in our system: speed, comprehension, and discourse organization. Vocabulary, which plays a crucial role in reading, is assessed separately and will be discussed later. Teachers can copy and paste or input a particular text to be tested. Alternatively, they can let the system do the job for them. By inputting relevant information such as difficulty, length, and topical category, the system can randomly choose a text in the database which meets the input conditions.

The semi-automatic generation of items is made possible through a corpus (i.e. a collection of texts) and text analysis tools. The difficulty level of texts is computed based on word frequency information derived from a corpus with the assumption that the word frequency of a text can be used as an indicator of text difficulty. It is assumed that the higher the average word frequency of a text, the easier the text. The topic of each text is manually encoded by a meta data tag.

The reading speed test is designed to train students to grasp the main idea of an article in the shortest possible time, while the reading comprehension test trains students to identify the writer's purpose and tone. The reading comprehension test also emphasizes the ability to quickly spot information relevant to a question and the ability to make inferences

based on a text.

To develop students' ability in organizing texts, we implement sentence and paragraph rearrangement tests. The order of sentences in a paragraph is scrambled and each sentence is marked with a number. Students are asked to rearrange these scrambled sentences back to its original order. There are also rearrangement tests at the text level where the order of paragraphs are misplaced and students are asked to restore it to the correct order. The sentence and paragraph rearrangement tests are integrative tests that combine reading comprehension with writing skills.

Through carefully-chosen texts from intermediate to advanced level, the reading speed tests, the reading comprehension tests, and the paragraph rearrangement tests help students develop good reading strategies such as skimming, scanning, glancing, guessing unfamiliar words using context, and identifying text organization. The student profiling component lists students' estimated reading speed, strength as well as weakness in reading. Through the profiling system, it is easy to check if a student has mastered a specific reading skill and how well he/she uses a particular reading strategy in an actual reading task.



Figure 1. Interface of a reading comprehension

test

Assessing Listening

There are three types of tests designed to assess students' listening ability in our system. The multiple choice questions are used to test students' grasp of the main idea or specific detail of a speech or conversation, while the cloze tests are used to test if students can identify and remember keywords and key phrases in a short conversation or speech. The multiple choice tests involve recognition skills. In contrast, the cloze tests involve production skills.

In addition to multiple choice questions and cloze tests, we have developed an on-line dictation program to assess listening ability. Dictation is generally believed to be an important approach to training and assessing listening ability. Involving spelling, punctuation, syntactic and semantic processing, dictation is recommended by many researchers (e.g. Oller (1971, 1979), Hughes (1989) and Weir (1990, 1993)) as a good method of integrative test of listening. Despite its usefulness, the effort required to score a dictation test prevents it from being widely used in examinations. Grading a dictation test is so tedious and time-consuming that most teachers feel daunted at the thought of it. We solve this problem by automating a dictation test using software tools and computational techniques. We use GoldWave, a shareware, to record a speech and edit waveform so that each sentence can be extracted and saved in MPEG III format. Once this step is done, all test administrator has to do is reduced to the following simple tasks: (1) uploading the sound file to a directory of the system (2) inputting the sentence and the number of times students can try, and (3) deciding the penalty of misspelling,

redundant words, and missing words. Using the algorithm of minimal edit distance, the program can detect the differences between the sentences input by students and those input by the test administrator. The system can automatically score students' performance based on previously input conditions of penalty. It can also perform data analysis by calculating the most common errors of all students as well as persistent errors made by a student.



Figure 2. The output of a dictation test

Assessing Writing

Our system has a component which facilitates the marking and revision processes by teachers and students in English composition classes. The system provides teachers with a tool to annotate different types of errors along with a database which records the number and types of errors students make. The system can perform data analysis and list the most common mistakes of a learner as well as of a class. When students login to the system, they can see teachers' annotations and comments and correct the mistakes by consulting on-line dictionaries and corpora linked to the system. If they are unable to correct the mistakes themselves, they can search the database and find how previous mistakes of the same type were corrected. Alternatively, they can follow the hyperlinks to

web pages that specifically address the same problem. The hyperlinks are created semi-automatically through existing relevant web pages.

The writing assessment tool integrates expert knowledge of a teacher with database and the Internet technology. It takes a process-based and tool-based approach to teaching and assessing writing. In other words, it emphasizes revision processes and the use of on-line resources. Rather than directly telling students how to correct mistakes, the system asks students to correct the identified mistakes themselves by learning from examples from their peers and on-line resources such as dictionaries and corpora. Our approach is justified because experienced English teachers are frustrated by correcting students' recurrent errors. Unless students actively construct and internalize the grammatical knowledge, they will repeat their mistakes regardless of teachers' correction. To make learning writing efficient and effective, we adopt an assessment method which takes into account the number of recurrent errors. Because recurrent errors incur low mark, students are more cautious when they write.



Figure 3. Teachers can annotate students' error types on-line.



Figure 4. Students' compositions annotated with error tags and hyperlinked to on-line resources.

Assessing Vocabulary

The vocabulary component can semi-automatically create item bank for English vocabulary tests. Currently we categorize test items into three levels: basic, intermediate, and advanced based on word frequency information derived from corpora. In the context of a proficiency test, the teacher can select a level and the system can randomly choose words which meet the input conditions. If the teacher wants to give an achievement test, he/she can input the words to be tested and the number of items to be tested, the system will automatically create a draft of vocabulary test using electronic dictionaries and a corpus. The load of the test administrator is thus reduced to editing the automatically generated items.



Figure 5. The vocabulary test in action

The data analysis module records the responses of each examinee to each item and score the test.



Figure 6. Test taker's responses to items

The data analysis module performs item analysis. It lists the item identifier in the item bank, the total examinees who were given this item, the total of examinees who got this item right, the item facility, the frequency information of the word tested, and the rank of the word in the corpus based on its frequency information.

Conclusion and Future Research

In this paper, we briefly describe our web-based assessment and student profiling system. The system can qualitatively and quantitatively assess students' ability in reading, listening, writing, and vocabulary using semi-automatic means. We are experimenting various data mining techniques to track and analyze students' learning processes and where possible, student's learning styles. The information will serve as input to the student modeling component of an intelligent tutoring system.

姓名	學號	聽力	口語	閱讀	寫作	總分
陳建人	9900001	85.0000	85.0000	85.0000	85.0000	85.0000
江明凡	9900002	85.0000	85.0000	85.0000	85.0000	85.0000
吳明輝	9900003	85.0000	85.0000	85.0000	85.0000	85.0000
吳明輝	9900004	85.0000	85.0000	85.0000	85.0000	85.0000
吳明輝	9900005	85.0000	85.0000	85.0000	85.0000	85.0000
吳明輝	9900006	85.0000	85.0000	85.0000	85.0000	85.0000
吳明輝	9900007	85.0000	85.0000	85.0000	85.0000	85.0000
吳明輝	9900008	85.0000	85.0000	85.0000	85.0000	85.0000
吳明輝	9900009	85.0000	85.0000	85.0000	85.0000	85.0000
吳明輝	9900010	85.0000	85.0000	85.0000	85.0000	85.0000

Figure 7. Score of individual language skill

Acknowledgements

The first author would like to thank Yu Shi-Jie 余世傑, Lin Zheng-Ru 林正儒, and Zhao Zheng-Ming 趙政銘, Lin Kui-Kuang 林桂光, Liang Jing-Shiu 梁菁秀, Shi Jia-Jun 施嘉峻 for implementing the system.

References

Brown, James Dean. (1997). Computers in Language Testing: Present Research and Some Future Directions. *Language Learning & Technology, Vol. 1, No. 1, pp. 44-59.*
<http://polyglot.cal.msu.edu/llt/vol1num1/brown/default.html>

Coniam, David. (1997) A Preliminary Inquiry Into Using Corpus Word Frequency Data in the

Automatic Generation of English Cloze Tests. *CALICO Journal, No 2-4, pp. 15- 33.*

Gao, Zhao-Ming. (2000) AWETS: An Automatic Web-Based English Testing System. *In Proceedings of the 8th Conference on Computers in Education/International Conference on Computer-Assisted Instruction ICCE/ICCAI, 2000, Vol. 1, pp. 28-634.*

Hughes, Arthur. (1989) Testing for Language Teachers. *Cambridge University Press.*

Jun, Da. (2000) Online Language Testing System. <http://www.bio.utexas.edu/jun/call/interactive/onlinetest.html>

McCormack, Colin, and Jones, David. (1998). *Building a Web-Based Education System. John Wiley.*

Roever, Carsten. (2001). Web-Based Language Testing. *Language Learning & Technology, Vol.5, No.2, pp. 84-94.*
<http://llt.msu.edu/vol5num2/roever/default.html>

Wilson, Eve. (1997) The Automatic Generation of CALL Exercises form General Corpora. *In Wichmann et al. (eds.) Teaching and Language Corpora, pp. 116 – 130. Longman.*