University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

A Similarity Threshold-based Tool for Generating and Assessing Essay Computer-Based Examinations

Olumide B. Longe
University of Ibadan, longeolumide@fulbrightmail.org

A Similarity Threshold-based Tool for Generating and Assessing Essay Computer-Based Examination

Longe Olumide Babatope
Fulbright Fellow
International Centre for Information Technology & Development
Southern University System
Southern University
Baton Rouge, LA, USA
longeolumide@fulbrightmail.org

Abstract
Most computer-based assessments (CBA) employ test generators that produce multiple choice questions usually without options. The limitations of these types of evaluation are that students can randomly select or guess answers with a 25% chance of choosing the right answer per question. The implication is that there is a one out of four probability that students can pass such examinations without understanding the contents taught in class, without studying for the examination and by just guessing answers. In light of the foregoing, the effectiveness of multiple choice and objective questions as a tool for evaluating students' mastery of subjects can be questioned. Unfortunately, most test generators do not have the capability for handling essay-based questions due to the fact that there are no rigid responses to essay examination questions. We attempted to bridge this gap by developing **EssayTest** - a tool that generates essay based questions and marks essay based examinations. Using JAVA, JDBC, MYSQL and other third party interface design tools, *EssayTest* employ similarity thresholds to match tokens in the answers supplied by teachers and responses from the students in an essay-based CBA as a way of scoring the examination. Preliminary tests showed very promising results.

Keywords: CBA, Essay, test, multiple-choice, questions, generators and examination.

Introduction

Testing is done in schools to determine if learners mastered what has been taught. Conventional examinations employ the use of paper answer booklets as a medium on which responses to paper-based questions can be answered. These are then collected at the end of each examination, marked and recoded to determine if the student should move to the next level or not or as a basis for course completion. Since their introduction in the early 1980s, personal computers (PCs) have seen great use in educational settings, from the development of computer based tutorials, computer-aided learning and computer based assessment. Test questions and results could be stored on computers days before it was printed out and so the number of people with access to the questions was reduced to just the teacher and the computer operator. In particular, testing has experienced a lot of improvements with the introduction of the internet. Now teachers teaching the same subject/course in different locations can come together

over the network and set the questions, sending the test over the network on the day of the exam (centralizing exam setting), thereby making it difficult for cheating to take place.

Even in light of all these developments, teachers are still actively involved with drawing up examination questions and could still make the assessment process porous by leaking questions either deliberately or inadvertently to students before the examination. Since teachers also mark essay examinations, there is the potential for undue favors and result manipulation to favor some students at the expense of others. In this paper, we present the development of an essay examination generator called *EssayTest*. This tool requires minimal external input in the generation and marking of questions and also provides a mechanism to store/record the results.

Related Works

Technology is playing an increasingly influential role in education globally. Computers and mobile phones are now being used to promote electronic learning and facilitate lectures across the globe in real-time mode (Sadiq, 2012). Multimedia facilities promote student engagement, interaction and collaborations in virtual learning environments. Technology is being used not only in administration and teaching but also for educational assessment (Dede, 2002). In some cases, conventional electronics with higher penetration, such as television and radio, are interlinked with the internet to reach learners in remote communities. For instance, the Kothmale Community Radio Internet employs this hybrid to provide educational opportunities in a rural community in Sri Lanka (Sally, 2008). The Indira Gandhi National Open University in India uses a combination of print, recorded audio and video, broadcast radio and conferencing technologies to reach learners (Mandanmohan, 2006). The distance learning program at the University of Ibadan, Nigeria also engages a mix of these technologies to reach learners on its Diamond FM station, the University Radio Station (UIDLC, 2012).

Existing standardized computer-based tests include the Scholastic Aptitude Test, or SAT, the Graduate Record Exam (GRE) to evaluate students applying for graduate degree programs, the Metropolitan Achievement Test (MAT), the California Achievement Test (CAT), the Comprehensive Test of Basic Skills, the Iowa Test of Basic Skills, the Preliminary Scholastic Aptitude Test (PSAT), taken in preparation for the SAT and used to select National Merit Scholarship winners, and the American College Test (ACT), an aptitude test taken in addition to or in place of the SAT.

While advocates of standardized tests maintain that test scores provide a valid measure of academic aptitude and contend that these examinations are impartial in comparing students from a variety of social and educational backgrounds (Schmitt and Dorans, 1988; Scholes and McCoy, 1998), critics argued that the tests do not account for differences in socioeconomic backgrounds and do not accurately assess the scholastic performance of students (Lorie and Graue, 1993; Willingham *et al*, 2000; Willingham et al, 1988; Young, 2004; Steele and Aronson, 1998). They also argued that emphasis on high test scores encourages teachers to focus only the material likely to be covered in the tests rather than provide a comprehensive education. Computer-Aided Learning (CAL) describes the use of technology in the teaching and learning process (Zwick, 2004; 2006). Test generators are computer programs that aid student's assessment process. The first test generator developed was really just a simple program that generated random numbers for each student and the questions corresponding to these numbers were printed out for the students to answer. The limitation of this system was that in major examinations, due to the number of students and the limited amount of questions, it is highly likely that more than one student would have the same questions to answer (Achim and Christophe, 2005).

New types of generators later came up that divide the students into groups and the students in each group were given completely different questions than those in all other groups. It became the responsibility of examiners to ensure that no two students in the same group sat down next to each other (Carlos and Abelardo, 2004). In this scenario, questions were still being generated by the teachers and answers to those questions were marked manually. An improvement on this was the development of automated test generators that allow teachers to set multiple choice questions that students can answer with automated result generation (Reggie *et al*, 2002). The difficulty of assessing students on essay type

examination led to the development of a new type of test generator that required teachers to submit lecture notes. The test generator then generates a table containing keywords (i.e. words that occur most in the note) and uses this to remove segments of the notes for students to fill in the missing parts. The challenge was that there were in lectures for which the keyword table contained mostly words that were not relevant to concepts that were taught in class, and as such any tests generated this way were not able to properly assess the student's understanding of the materials presented in class. These systems are fraught with so many challenges that in most educational settings, essay-based tests using computers have been completely abandoned and only multiple-choice computer assessment are in use.

Research Direction

Most computer-based assessments (CBA) employ test generators that produce multiple choice questions usually without options. The limitations of these types of evaluation are that students can randomly select or guess answers with a 25% chance of choosing the right answer per question. The implication is that there is a one out of four probability that students can pass such examinations without understanding the contents taught in class; without studying for the examination and by just guessing answers. Multiple choice testing cannot provide in all cases evidence that students have learned and therefore is not an effective way of testing students' ability in some courses.

To overcome the limitations of the keyword-based system, we proposed and develop an essay type test generator that allows teachers input "likely questions" and answers into a database. Questions are then randomly selected from these pools and assigned randomly to students. Answers in response to these questions by the students are then compared to pre-recorded answers in the database and students are graded. The system will randomly assign questions to students in such a way that no two students will have the same question at any point in time. To mark the paper the answer submitted would be broken up into tokens and this would be tested against the answer given by tutors to see if they mean the same thing. The use of a database would be employed to save all the data and facilitate easy comparison, recording and updating.

EssayTest System Overview

Our intention is to automate the whole essay-based testing process such as the setting questions, grading answers, storing and displaying results. The system allows the tutor to input questions before the examination and these are saved in the database. Out of all these questions previously submitted into the database by the tutor via the lecturer's panel / side of the system, random questions are generated for each student sitting for the examination. If enough questions have been inputted into the database, no two (2) students in the entire hall would be answering the same question at the same time. After the exam the answers are marked automatically and graded and stored in the database. To re-mark scripts the results of the student are recalled from the database with the click of a button.

The main users of this system would be:
- Students
- Professors/Tutors/Lecturers
- Administrators

Students would sit for exams via the system and their scripts would be graded by the system. The lecturers would also use the system to set questions, view result and search for scripts. The administrator is an individual who will be placed in charge of the testing system. The job of the administrator is to regulate the system, manage users in such a way that they would not interfere with each other, check to see that questions set by lecturers are up to standard, view results, search for scripts and set exam time and duration. There are scenarios which provide a more specific view of the different functions that must be implemented to perform many of the general functions mentioned above. The scenario is listed below.

3

"A Similarity Threshold-based Tool for Generating and Assessing Essay Computer-Based Examination," Longe Olumide Babatope.
Library Philosophy and Practice 2012

System Design

The system design was divided into two phases:
1. Logical Design
2. Physical Design

Logical Design

A logical data flow diagram shows the flow of data through a transaction processing system without regard to the time period when the data flows or the processing procedures occur. Here I designed the software logically, using process modeling by Data Flow Diagram (DFD) and Entity Relation diagram (ERD) technique.

Physical Design: A user-friendly interface was developed for the *EssayTest* Generator using Java Programming.

Use Case
The Use Case Diagram is a UML Diagram that is used to show the actors in a given system and the activities they perform. The actors are as follows:
1. Students
2. Lecturer

List of Use Cases
1. Get Question
2. Answer Question
3. Mark Answers
4. Input Questions
5. Input Answers
6. View Result
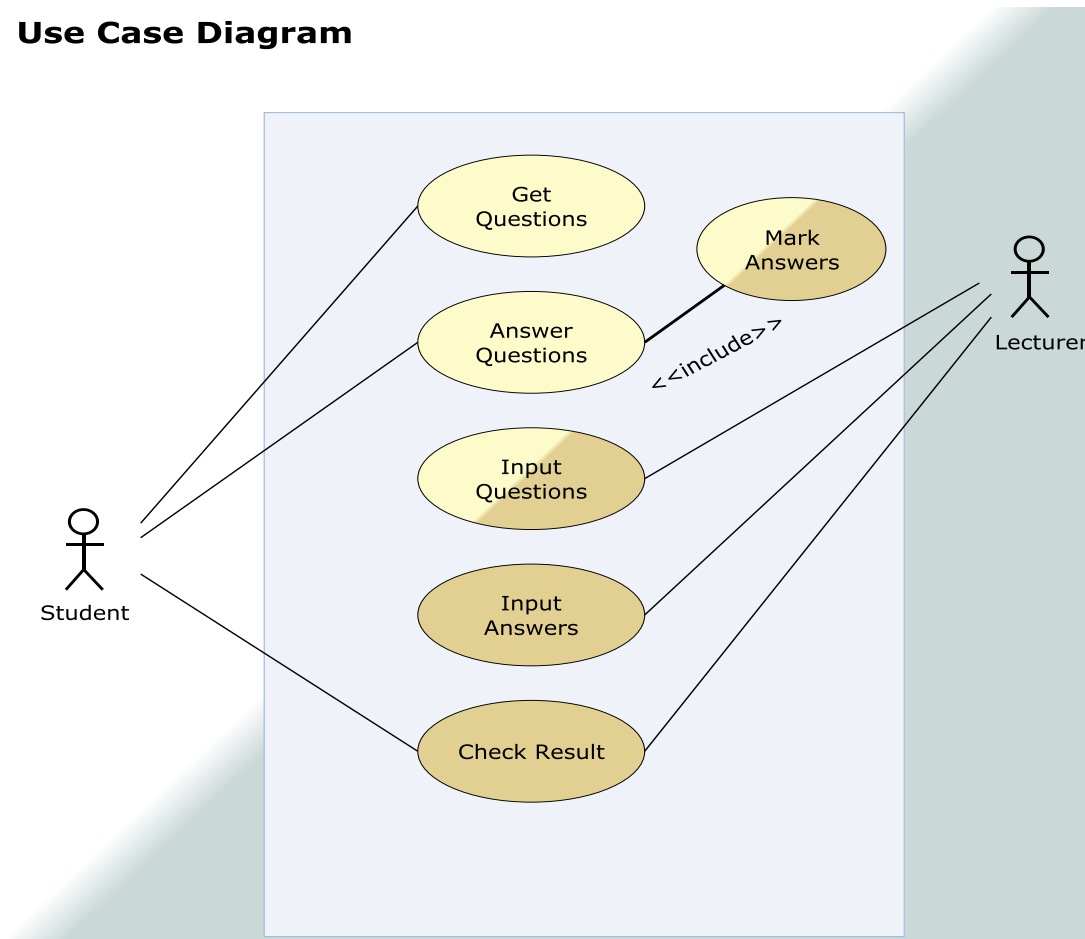
Use Case Diagram

Fig. 2: Use Case Diagram

System Implementation

The system consists of two usage platforms, the lecturer's platform and the student's. At the lecturer's end, there are seven (7) sections, namely login, add questions, edit questions, view results, remark scripts, modify login, course details, while on the student side there are basically only three (3) sections – login, answer questions and results.

Login Section

This is the first page that lecturer will see when they start the application. This section enables the entire application be secured, as people without the correct passwords are not allowed to access any other part of the application. It also ensures that lecturers are only able to access their own courses and no one else's.

Fig. 3: Lecturer's Login Page

Home
If the user has successfully entered correct login details, he is taken to the home page, which contains shortcuts to other sections of the application.

Fig. 4: Lecturer's Home Page

Input Questions
On this page the lecturer can enter questions, answers to those questions and the respective marks of those questions, which will be saved in the database.
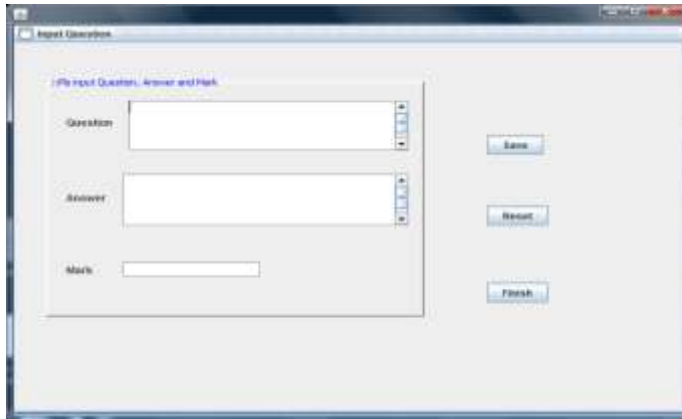
Fig. 5: Input Questions Page

Review Questions

Here the lecturer can view the questions currently in the database and can modify the question, its answer or the mark associated to it and save this updated version in the database, or he can delete the question from the database.
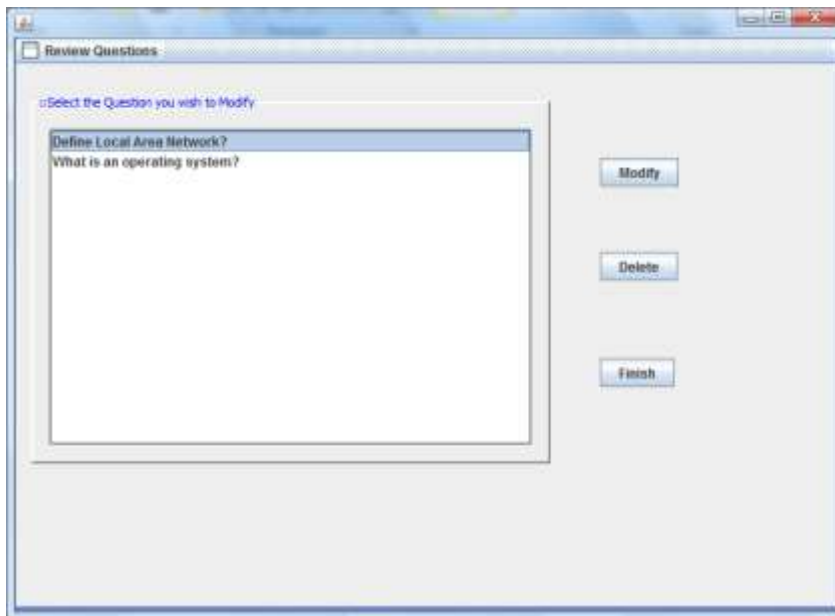

Fig. 6: Review Questions Page

Check Result

Here the lecturer can view the results of the students who took the course that year and can also print this result.
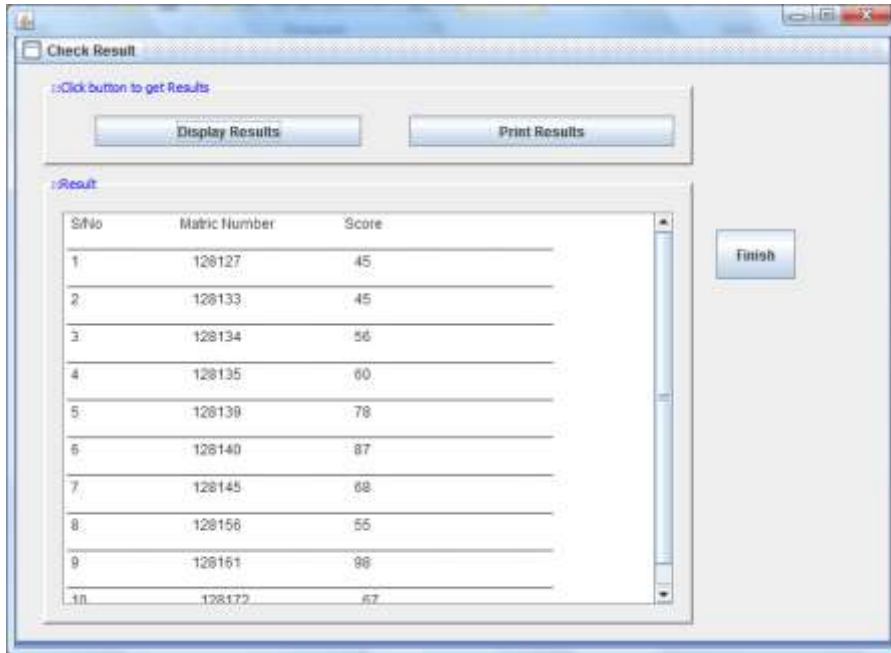
Fig. 7: Check Result Page

Student's Login

On the day of the exam students can sit for their exams via the student's side of the application. To do so they first have to login with their Student ID number to ensure that only registered students are permitted to sit for the exam.
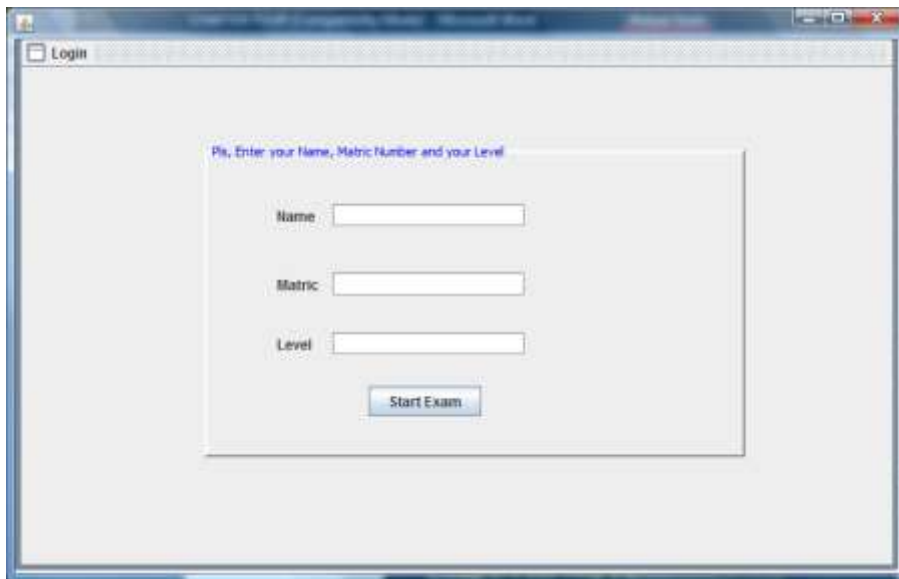

Fig. 8: Student's Login Page

Exam Page

On login, questions are randomly generated and given to the students until either the allocated time for the exam elapses or the student answers enough questions.
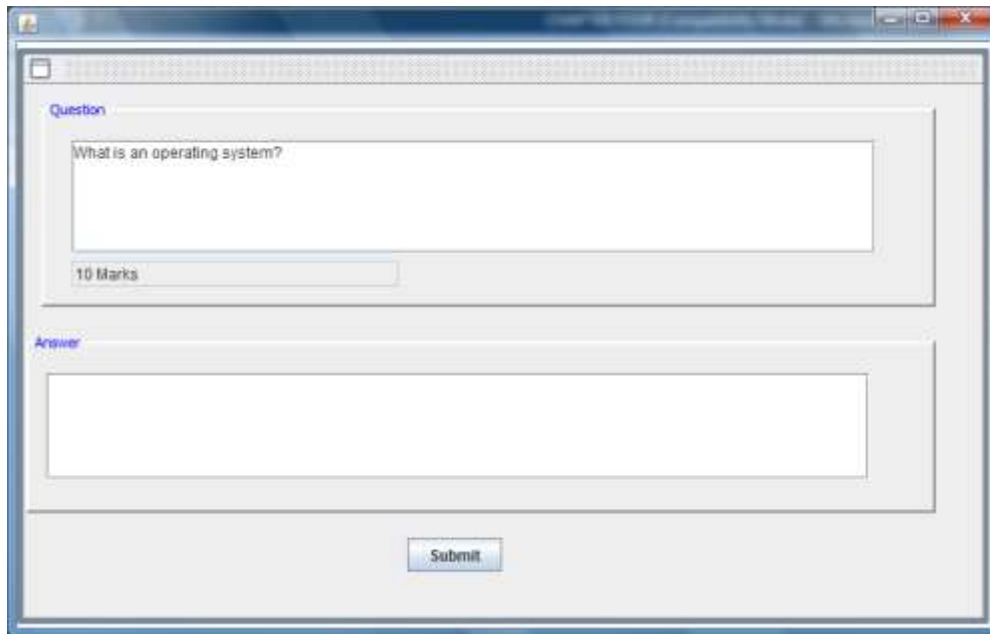
Fig. 9: Exam Page

Conclusion and Future Works

Most computer-based assessments (CBA) employ test generators that produce multiple choice questions, usually without options. The limitations of these types of evaluations are that students can pass them without possibly having a mastery of the concepts taught. We developed *EssayTest*, an automated test generator for essay examination, as a solution to the inadequacies associated with multiple choice computer-based examinations. Future work will seek to increase and improve system functionalities by providing components for biometrics authentication for test takers and the ability for the system to upload graphics or answers that require the student to draw.

End Notes /Appreciation

The author appreciates the efforts and collaboration of Mr. Abiodun Ajayi for his input into programming the interface.

References

Achim, R. and Christophe, B. (2005): New trends and technologies in computer-aided learning. FIP TC10 Working Conference: EduTech 2005, October 20-21, 2005, Perth, Australia. https://library.villanova.edu/Find/Record/986996

Carlos, D. and Abelardo. P. (2004): computer-aided design meets computer-aided learning. IFIP 18th World Computer Congress ; TC10/WG10.5 EduTech Workshop, 22-27 August 2004, Toulouse, France. www.informatik.uni-trier.de/~ley/db/indices/a

Dede, C. (2002). Vignettes about the future of learning technologies. In Visions 2020: Transforming education and training through advanced technologies. Washington, DC: U.S. Department of Commerce. http://www.technology.gov/reports/TechPolicy/2020Visions.pdf
.
Reggie, K., Jimmy, C., Weijia, J., Anthony, F. and Ronnie, C. (2002): Web-based learning: men & machines. Proceedings of the First International ... of the First International Conference on Web-

Based Learning in China (ICWL 2002) .
http://isbndb.com/d/publisher/world_scientific_publishing_co.html?start

Sadiq, F.I (2012). eCollaboration for Tertiary Education Using Mobile Systems. Computing, Information Systems & Development Informatics Journal. Vol 3, No.1. pp 5-9

Schmitt, A. P. & Dorans, N. J. (1988). Differential item functioning for minority examinees on the SAT (ETS Research Report 88-32). Princeton, NJ: Educational Testing Service.

Scholes, R. J., & McCoy, T. R. (1998, April). The effects of type, length, and content of test preparation activities on ACT assessment scores. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Shepard, L.A., & Graue, M.E. (1993). "The morass of school readiness   testing: Research on test use and test validity." In B. Spodek (Ed.), *Handbook of Research on the Education of Young Children.* New York: Teachers College Press.)

Steele, C. M., & Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In C. Jencks & M. Phillips (Eds.), The BlackWhite test score gap (pp. 401-427). Washington, DC: Brookings Institution Press.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2000). Grades and test scores: Accounting for observed differences (ETS Research Report 00-15). Princeton, NJ: Educational Testing Service.

Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., & Powers, D. E. (1988). Testing handicapped people. Boston: Allyn and Bacon, Inc.

Young, J. W. (2004). Differential validity and prediction: Race and sex differences in college admissions testing. In Zwick, R. (Ed.), Rethinking the SAT: The Future of Standardized Testing in University Admissions. pp. 289-301. New York: RoutledgeFalmer.

Zwick, R. (2004). Is the SAT a "wealth test?" The link between educational achievement and socioeconomic status. In R. Zwick (ed.), Rethinking the SAT: The Future of Standardized Testing in University Admissions, pp. 203-216. New York: RoutledgeFalmer.

Zwick, R. (2006). Higher education admissions testing. In R. L. Brennan (Ed.), Educational Measurement (4th ed.), pp. 647-679 Westport, CT: American Council on Education/Praeger.

EDITOR: The following references were in the bibliography but are not referred to in the article:

Madanmohan, R. (2006) The nature of the information society: A developing world perspective. http://www.itu.int/osg/spu/visions/papers/developingpaper.pdf

Sally D. B. (2008) Food & Agriculture Organisation, Rome MDE Programme, Athabasca University, Canada. http://www.irrodl.org/index.php/irrodl/article/view/563/1038

Sawyer, R. L. (1985). Using demographic information in predicting college freshman grades. (ACT Research Report No. 87) Iowa City: ACT, Inc.

UIDLC (2012). University of Ibadan, Nigeria Distance Learning Centre. http://www.dlc.ui.edu.ng/sub-degree--diplomalife-long/admission-prospective-applicants