WEB-BASED ADAPTIVE TESTING SYSTEM (WATS) FOR CLASSIFYING STUDENTS ACADEMIC ABILITY

Jaemu LEE
Busan National University of Education, SOUTH KOREA

Sanghoon PARK
Northwestern State University of Louisiana, U.S.A

Kwangho KIM
Busan National University of Education, SOUTH KOREA

ABSTRACT

Computer Adaptive Testing (CAT) has been highlighted as a promising assessment method to fulfill two testing purposes: estimating student academic ability and classifying student academic level. In this paper, we introduced the Web-based Adaptive Testing System (WATS) developed to support a cost effective assessment for classifying students' ability into different academic levels. Instead of using a traditional paper and pencil test, the WATS is expected to serve as an alternate method to promptly diagnosis and identify underachieving students through Web-based testing. The WATS can also help provide students with appropriate learning contents and necessary academic support in time. In this paper, theoretical background and structure of WATS, item construction process based upon item response theory, and user interfaces of WATS were discussed.

Keywords: Computer based testing, Assessment, Adaptive testing, Web based testing system.

INTRODUCTION

The rapid advancement of computer technology in education has radically changed the school structure and curriculum as well as teaching practices in the classroom. It also has significantly influenced the administration of assessment systems in schools. As a practical alternative to traditional paper-and-pencil testing, computerized testing was introduced a decade ago and has been a viable option for many schools and institutions seeking efficient ways to administer assessment tests (Meijer & Nering, 1999; Ruiz, Fitz, Lewis, & Reidy, 1995; Weiss, 1983). Recently, the use of computerized tests has been spotlighted again with an announcement of U.S. Department of Education (2011) that one of the two major assessment consortia will be using a Computerized Adaptive Testing.

Computerized Adaptive Testing (CAT) is a computer-based method for constructing and delivering individualized testing instruments. The key advantage of CAT is the test item selection algorithm, which customizes the test items by altering them to best fit each examinee's ability level based on his provisional ability estimates (Moore, Galindo, & Dodd, 2012; Han, 2009; Wainer, et al., 1990).

The development of CAT is grounded in computer technology and Item Response Theory (IRT) (Leung, Chang, & Hau, 2003; Meijer & Nering, 1999). The introduction of IRT has provided a method in which we can ascertain an examinee's proficiency from his/her performance on a set of items that are different depending on individual student's academic ability (Davis, 2004; Kingsbury & Houser, 1983). CAT also has provided a customized and precise assessment for each examinee. The premise of CAT is based on the fact that too easy items or too difficult items contribute little to the information about an examinee's trait level (Lilly, Barker, & Britton, 2004). By eliminating items of inappropriate difficulty, CAT can shorten testing time, increase measurement precision, and reduce measurement error due to boredom, frustration, and guessing. The number of applications of CAT is growing quickly and psychometric research on adaptive testing has received widespread attention. Since the initial CAT research performed in the 1970s and 1980s, studies on the implantation of CAT have been grown, and several tests have adapted an operational CAT version such as the Graduate Record Examination (GRE) and the computerized placement test (Lord, 1971, 1977; Weiss, 1982, 1983). The use of CAT has been increasing and replacing traditional Pencil and Paper tests in education and training settings. Usually this replacement is associated with the need for higher efficiency when assessing large number of students, for example, in online training. GMAT, TOEFL, and Microsoft Certificate Test are all evidence of this trend.

Applications of CAT serve for two purposes (Eggen & Straetmans, 2000; Ruiz, Fitz, Lewis, & Reidy, 1995). First, it is used for ability estimation of students. In learning or a training context, the goal of evaluation is to rank students according to the level of ability. For ability testing, CAT begins with the items of moderate difficulty. If an examinee selects an incorrect answer, an easier item than the previous one will appear next. If an examinee selects the correct one, the item pool automatically selects the more challenging item. Since the computer selects items on the basis of previous responses, an examinee's ability can be evaluated continually (Meijer & Nering, 1999; Weiss, 1983). The second purpose of CAT applications is to classify students' level. Traditionally CAT aims at the efficient estimation of an examinee's ability. However, it also has shown to be a useful approach to classification problems. Weiss (1983) described CATs for situations where the main interest is not only in estimating the ability of an examinee, but to classify the examinee into one of two categories, e.g., pass-fail, master/non-master (Chen, Wigand, & Nilan, 1999; Eggen & Straetmans, 2000; Ruiz, Fitz, Lewis, & Reidy, 1995). Since the computer algorithm selects items on the basis of previous responses, an examinee's ability is evaluated continually and ranked according to the score.

With the benefits and promises of CAT, however, there are not many adaptive tests available in K-12 school settings. Chang (2012) pointed out that the cost effectiveness of hardware and network design that school can afford is an important issue when utilizing CAT in K-12 classrooms. He further added the function of cognitive diagnoses must be incorporated into the CAT using an item selection algorithm. In a traditional classroom testing environment, an identical set of test items is given to students at the same time in the same place to diagnosis each student's cognitive level without considering individual differences.

As Underachieving students and high achieving students are given the identical set of test items, the items consequently fail to classify students' academic ability into underachiever and high achiever.

Teachers, as an evaluator, spen considerable amount of time to develop, evaluate, and analyze test items to use for classifying students' level. To meet these challenges, the need of developing an Web based adaptive testing system has been increased. Teachers need an adaptive testing system to effectively manage test items they developed, to control the number and difficulty levels of the items, and to retrieve student's test information in order to provide just in time feedback and academic support to underachieving students.

In this paper, a case of developing a cost effective Web-based Adaptive Testing System (WATS) is discussed. The WATS is intended to

> ➢ help teachers create test items online as well as save item information,
> ➢ help teachers retrieve adaptive test items with consideration of individual student's ability, and
> ➢ help teachers manage the length of test, item difficulty, and test results to diagnosis a student's academic ability in K-12 settings.

DEVELOPMENT OF WATS (WEB-BASED ADAPTIVE TESTING SYSTEM)

Purpose of the WATS
The purpose of the WATS is twofold. First, we expect the WATS to diagnosis underachieving students by adaptively responding to students' given answers. Second, the WATS is designed to provide test reports that are automatically sorted by specific criteria to help teachers identify underachieving students. The following design requirements were considered before developing the WATS.

First, WATS test item pools must be structured according to students' grade levels, test item difficulty, and test factors. All three sets of information should be documented in the testing structure to adaptively present a test item with an appropriate difficulty level to each student. Second, the WATS should be designed to present each student a set of questions that are appropriate to his level of academic ability.

Third, testing time should vary depending on a student's academic ability. Fourth, test results must be available once a student completes the test. Finally, Web based access to the test score database should be available for teachers to examine test result and academic progress for each student.

Work flow of the WATS. Flowchart illustrated in Figure: 1 shows how a student processes the entire WATS. The WATS is not designed to present test items in a linear format. Instead, test items are dynamically selected and retrieved for each student based on his answers during the test.

A test session begins with a randomly generated test item with an average difficulty level. If a student answers to the item correctly, the probability of a correct estimate of his academic ability is increased. The process continues until the estimation error rate is acceptable based on a pre-defined error rate level.Structure of WATS testing system. The complete structure of test system is illustrated in Figure: 2. The system consists of an administrator mode, an instructor mode, and a student mode. The test mode is determined by login information and users are guided into three different modes depending on each role.
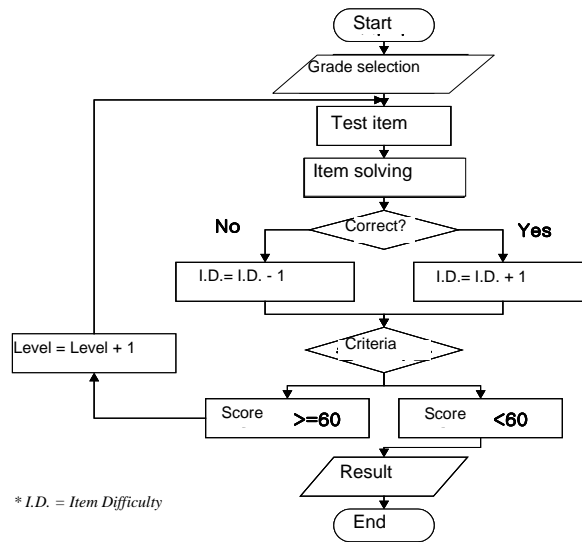
Figure : 1
Work flow of adaptive testing system

In particular, administrator mode contains "test item management", "student information", and "instructor information". Instructor mode includes "student reports and "test trend analysis. Student mode provides several menus such as online diagnosis and my report".
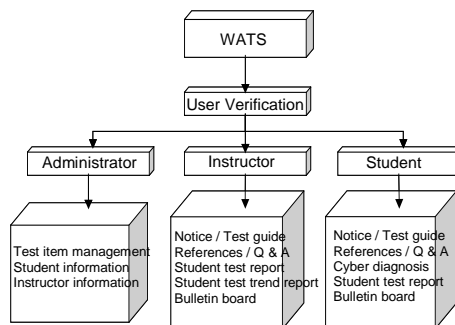

Figure: 2
Testing system structure: Administrator, Instructor, and Student mode

Flow chart of WATS system. The flow chart suggested by the developer of the WATS is illustrated in Figure 3. A student can access to the system through a user identification process. Once the student selects a grade level, he is guided to the actual test mode. Test items are retrieved from the item database.

The answer of the student is compared with the correct answer information saved in the test item database. The result of the entire test is saved in a report database after confirming with the student information database. Teachers can access to the test report database to check students' scores.
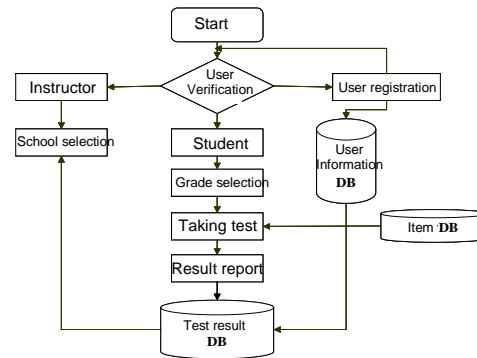
Figure: 3
Test system flowchart

When a student selects the appropriate grade level, WATS session starts with a random question of average difficulty. If the student answers the question correctly, the probability of correctly estimating his ability is increased. With the ability estimate increased, the students are assumed to be able to answer a more difficult question.

Thus, a more challenging question follows. Conversely, if a student makes an incorrect answer, the estimate of his ability is decreased and consequently, an easier question that is suitable for this new lower estimate is then presented.

OPERATION OF WATS

In this system, Windows 2000 professional operating system, IIS 5.0 web server, MS-SQL 2000 for databases, ASP utilizing ADO server components were implemented.

Student Mode
The main menu of student mode is activated when a student logs in to the system. The main menu screen consists of two sub menus, "Test" and "Report". A student can choose his grade level and take a test by clicking on the "Test" menu.

Final scores are available upon a student's request using the "Report" menu. Also further information of previous tests results can be retrieved from the test score database. Figure: 4 shows one of the sub menus, "Test".

The reason the system includes "grade selection" is to help underachieving students correctly select an ability level when they perceive that their ability is lower than the actual grade level.

Figure: 5 shows actual testing screen where the first test item is randomly presented. The sequence of the next items is determined based on the answer for the first item. If a student answers the first item correctly, the next item is retrieved from the item pool with higher difficulty.
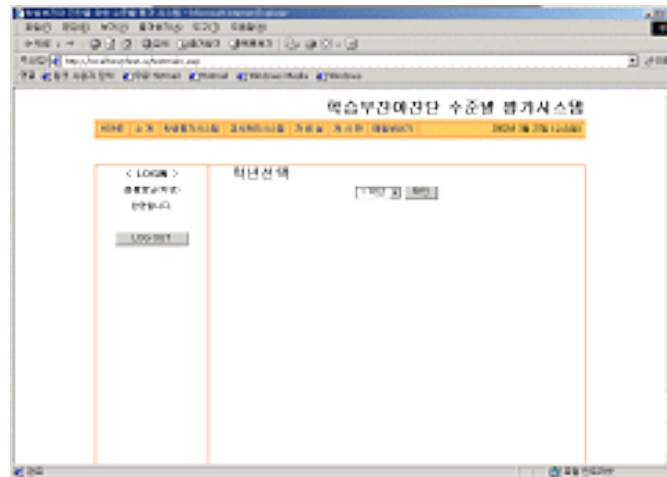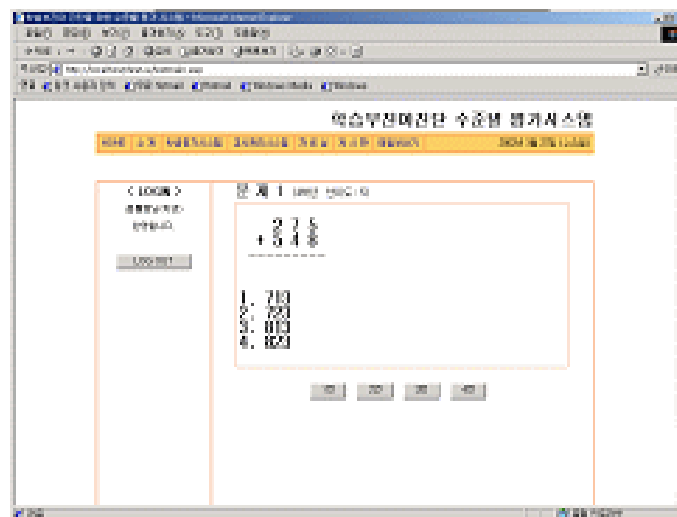
29

Figure: 4
Grade selection screen

Figure: 5
First test item screen (Item difficulty 5)

If the answer was incorrect, a less difficult item is retrieved from the item pool with lower difficulty. The basis of the item retrieval is on IRT and CAT. Figure 6 presents the case when a student answered an item incorrectly.

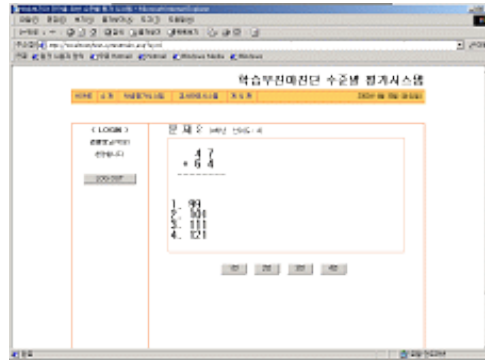A less challenging item (item difficult level 4) is retrieved using the algorithm of the IRT principle.

Figure: 6
Second test item screen (Item difficulty 4)

Figure 7 shows an item with lower difficulty level (item difficult level 2) when the student made one more incorrect answer.
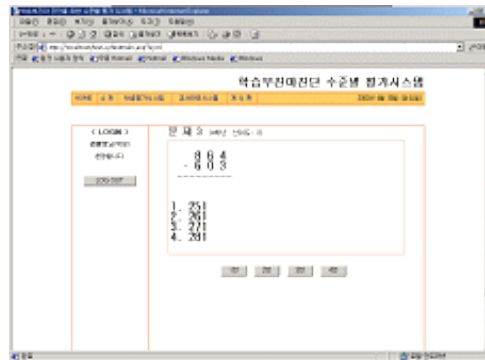

Figure: 7
Third test item screen (Item difficulty 2)

Figure 8 describes the item selection process applied in WATS system. The second test item is retrieved according to the answer the student made. For example, the second item (Item difficulty 4) appears when a student made an incorrect answer for the first item (Item difficulty 5). In same way, if the student made an incorrect answer for the second item (Item difficulty 4), the third item (Item difficulty 2) is given to the student.
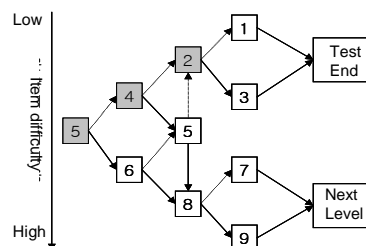

Figure: 8
Item selection process

This process ends when the student scores lower than the criteria score pre-determined to diagnosis underachieving students. If a student's score is classified to be in the underachieving level, the test stops running, and the system shows the test result.

Figure 9 shows the test results of student who took the test. It shows that students achieved a total of less than 60. Since 60 points was a pre-determined score criterion to diagnosis underachieving student, he is classified as an underachieving student. Also additional information is presented to specify which topics or concepts the student has most troubles with. In this case, the student was having a trouble with the concept of [Adding numbers], therefore a teacher is able to prepare extra study materials to help the student learn how successfully learn [Adding numbers]. On the contrary, if a student scores more than decision making criteria, he is guided to a more challenging step.

Instructor Mode

The main menu of instructor mode is activated when a teacher logs in to the system. The main menu consists of three sub menus such as "student information management", "testing trends review", "my information management .
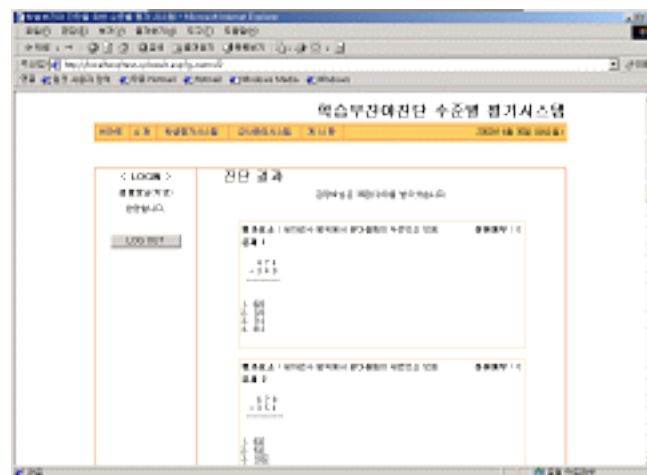

Figure: 9
Test result screen

"Student information management" menu allows a teacher to review student's information containing grade, name, recent testing date, and test result as shown in Figure 10. Test scores are listed according to the test date with total number of correct answers and classification results.

"Testing trend review" displays the items which the student perceived most difficult.

This information is useful for teachers to prepare extra materials for underachieving student.
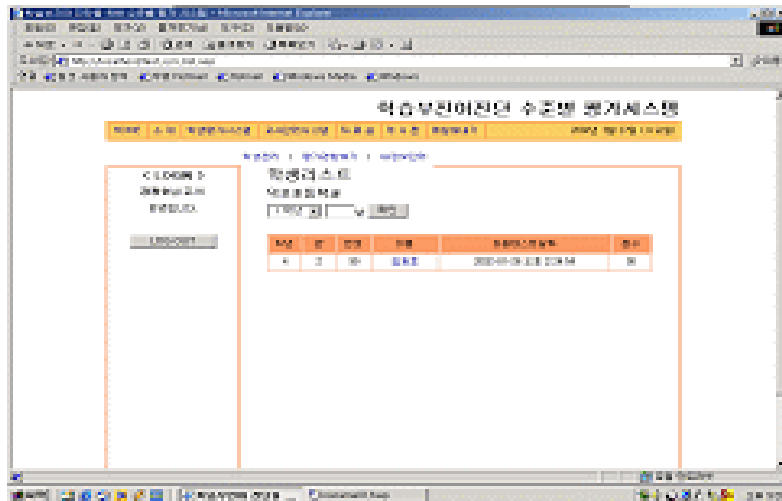
Figure: 10
Student information management screen

CONCLUSION

In this paper, the structure and interface of WATS were described. WATS controls the length, type, and difficulty of test items by adaptively responding to a student's ability and further classifying students into different levels in an efficient way. Teachers can create test items easily and save them on the database, a component of WATS. The expected effects of using WATS are following.

First, the minimum number of items can be used to correctly diagnosis a student's level by presenting items with controlled difficulty levels based on the student's ability. Also timely reporting of test scores is supported by efficiently controlling the time taken to develop, operate, grade, and report the test. Unlike a traditional paper-and-pencil test, WATS will help teachers spend less time and effort for assessing students' ability.

Second, teachers can look up the statistical data easily by having the system report and retrieve the test results sorted by individual student, test scores, or testing dates. Therefore, appropriate feedback can be easily prepared for students who have a trouble understanding a certain topic or a concept.

Third, the speedy process of diagnosing underachieving students can be completed utilizing a database where student's information is saved and analyzed. Further study is needed to investigate the effects of WATS implementation in a school setting by comparing it with paper-and-pencil tests that are extensively used in schools to diagnosis students' academic level.

AUTHORS' NOTE: Correspondence concerning this article should be addressed to Dr. Sanghoon Park, Educational Technology program, College of Education and Human Development, Teacher Education Center, Northwestern State University of Louisiana, Natchitoches, LA, 71497. Phone: 318-357-5774. Fax: 318-357-6275. E-mail: parks@nsula.edu

BIODATA and CONTACT ADDRESSES of AUTHORS

Jaemu LEE, Ph.D. has been a professor in the Computer Education Department at the Busan National University of Education in Korea since 1987. He obtained his Ph.D. in Computer Science from the Hongik University at 1994 in Korea. He was vice president of the Korea Information Education Association from March of 2005 to February of 2007.. His research interests include intelligent tutoring systems, adaptive learning systems, educational ontology, and instruction methods for the Computer.

Computer Education Department
Busan National University of Education in Korea
Busan, South Korea
Tel: 82-51-500-7322, Fax: 82-51-500-7321
Email: jmlee@bnue.ac.kr
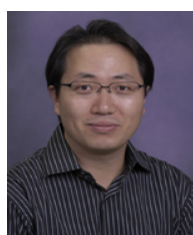
Sanghoon PARK, Ph.D. is currently serving as an associate professor and program coordinator of Educational Technology program at Northwestern State University. He has completed his Ph.D. in Instructional Systems at Florida State University with an emphasis in Instructional Design and Interactive Learning Technologies.

Sanghoon PARK, Ph.D.
Department of Educational Technology and Leadership
College of Education and Human Development
Northwestern State University of Louisiana
Natchitoches, LA 71457. U.S.A.
Tell: 1-318-357-5774, Fax: 1-318-357-6275
Email: parks@nsula.edu

Kwangho KIM is working as a teacher at the Keummyoung elementary school in Busan, Korea. He received his Masters degree from the Department of Computer Educational at Busan National University of Education in Korea. His research interests include intelligent tutoring systems, adaptive learning systems and testing system.

Kwangho KIM
Keummyoung elementary school
Busan, South Korea
Tel: 82-10-9396-3274, Fax: 82-51-500-7321
Email: xcdr74v@nate.com

REFERENCES

Chang, H-H. (2012). *Making computerized adaptive testing diagnostic tools for schools*. In R. W. Lissitz & H. Jiao (Ed.), Computers and their impact on state assessment: Recent history and predictions for the future (pp. 195-226), Information Age Publisher.

Chen, H., Wigand, R. T., & Nilan, M. (1999). Flow experiences on the web. *Computers in uman Behavior, 15*(5), 585-608.

Davis, L., L. (2004). Strategies for Controlling Item Exposure in Computerized Adaptive Testing With the Generalized Partial Credit Model. *Applied Psychological Measurement, 28*, 165-185.

Eggen, T. J., & Straetmans, G. L. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement, 60*, 713-734.

Han, K. T. (2009). A gradual maximum information ratio approach to item selection in computerized adaptive testing 2009 GMAC conference on computerized adaptive testing.

Kingsbury, G., & Houser, R. (1983). Assessing the utility of item response models: Computer adaptive testing. *Educational Measurement: Issues and Practice, 1*(12), 21-27.

Leung, C. K., Chang, H. H., & Hau, K. T. (2003). Computerized Adaptive Testing: A Comparison of Three Content Balancing Methods. *The Journal of Technology, Learning, and Assessment, 2*, 3-15.

Lilly, M., Barker, T., & Britton, C. (2004). The development and evaluation of a software prototype for computer-adaptive testing. *Computers & Education, 43*, 109-123.

Lord, F. M. (1971). A theoretical study of two-stage testing. *Psychometrika, 36*, 227-242.

Lord, F. M. (1977). Abroad-range tailored test of verbal ability. *Applied Psychological Measurement, 1*, 95-100.

Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement, 23*, 187-194.

Moore, E. L., Galindo, J. L., & Dodd, B. G. (2012). Balancing flexible constraints and measurement precision in computerized adaptive Testing. *Educational and Psychological Measurement, 72*(4) 629–648

Ruiz, B., Fitz, P., Lewis, C., & Reidy, C. (1995). Computer-adaptive testing: A new breed of assessment. *Journal of American Dietetic Association, 95*, 1326-1327.

U.S. Department of Education (2011, October 4). Smarter Balanced Assessment Consortium Core Components. Retrieved from http://www.k12.wa.us/SMARTER/pubdocs/SBACSummary2010.pdf

Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F.,Mislevy, R. J., Steinberg, L., & Thissen, D. (1990). *Computerized adaptive testing: A primer*. Hillsdale NJ: Erlbaum.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement, 4*, 273-285.

Weiss, D. J. (1983). *New horizons in testing: Latent trait test theory and computerized adaptive testing*. New York, NY: Academic Press.