DOCUMENT RESUME

ED 358 154   TM 019 953

AUTHOR   Wise, Steven L.; And Others
TITLE   The Role of Anxiety in Examinee Preference for
              Self-Adapted Testing.
PUB DATE   Apr 93
NOTE   18p.; Paper presented at the Annual Meeting of the
              American Educational Research Association (Atlanta,
              GA, April 12-16, 1993).
PUB TYPE   Reports - Research/Technical (143) --
              Speeches/Conference Papers (150)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   Ability; *Adaptive Testing; Algebra; *College
              Students; Comparative Testing; *Computer Assisted
              Testing; Higher Education; Mathematics Achievement;
              *Mathematics Anxiety; Mathematics Tests; Pretests
              Posttests; *Student Attitudes; Test Anxiety; Test
              Format; Test Selection
IDENTIFIERS   *Preference Patterns; *Self Adapted Testing

ABSTRACT
              This study assessed whether providing examinees with
a choice between computerized adaptive testing (CAT) and
self-adaptive testing (SAT) affects test performance in comparison
with being assigned a CAT or SAT, and evaluated variables influencing
examinee choice of either test form. The relative influences of test
type and test choice on examinee anxiety were also examined. Subjects
were 244 undergraduate and 133 graduate students from a large
midwestern university. Students were randomly assigned to SAT, CAT,
and choice conditions for an algebra test. Test-related anxiety was
assessed with a paper-and-pencil measure in pretests and posttests.
It was found that, for students with high mathematics anxiety,
providing a choice between CAT and SAT led to significantly higher
mean proficiency estimates, lending support to the hypothesis that
examinees can cope with a stressful situation more effectively if
they feel that they have some control over the source of the stress.
Expected differences in estimated proficiency and posttest state
anxiety between CAT and SAT conditions were not found, but a strong
relationship was seen between examinee test type choice and
mathematics anxiety level. Higher anxiety examinees have a greater
preference for the control provided by SAT. Six tables and two graphs
summarize findings. (SLD)

The Role of Anxiety in Examinee Preference for

Self-Adapted Testing

Steven L. Wise

Linda L. Roos

Barbara S. Plake

Lori J. Nebelsick-Gullett

University of Nebraska-Lincoln

The Role of Anxiety in Examinee Preference for Self-Adapted Testing

Computerized adaptive testing (CAT) is an increasingly popular application of item response theory (IRT). Using a pool of calibrated test items, a computer algorithm is employed to match the difficulties of the items administered to the proficiency level of each examinee. Because each examinee receives a CAT that is tailored to his/her proficiency level, substantially fewer items are needed per examinee in order to attain the same level of measurement precision as with a conventional test. Efficient testing is the primary advantage of CAT.

Efficiency is not, however, the only benefit that can be gained from the use of IRT in computer-based testing. Several years ago, Rocklin and O'Donnell (1987) explored an innovative application of IRT in computerized testing, termed *self-adapted* testing, in which the difficulty levels of the items administered are chosen by the examinee, rather than by a computer algorithm (as in a CAT). They found that examinees who received a self-adapted test (SAT) scored significantly higher (in terms of IRT-based proficiency estimate) than examinees receiving a conventional computerized test. Rocklin and O'Donnell interpreted the higher scores on the SAT as an indication that examinees were able to make effective and strategic choices among the items.

Subsequent research studies have explicitly compared SAT and CAT. Rocklin and O'Donnell (1991) found that, using a SAT, examinee test performance was less influenced by anxiety than when a CAT was used. Wise, Plake, Johnson, and Roos (1992) compared the test performances of examinees who were randomly assigned to take either a SAT or a CAT. They found that, relative to the CAT, examinees taking the SAT showed (a)

significantly higher mean proficiency estimates and (b) significantly lower post-test state anxiety. Using a sample of junior high school students, Vispoel and Coffman (in press) compared SAT and CAT versions of a music listening test, finding that (a) the SAT yielded higher mean estimated proficiency and (b) performance on the SAT was less influenced by test anxiety.

A recent study by Roos, Plake, and Wise (1992) investigated the importance of item feedback (which was used in the Rocklin and O'Donnell, Wise et al., and Vispoel and Coffman studies) in self-adapted testing. Roos et al. compared SAT and CAT, with item feedback either present or absent. It was found that the self-adapted test yielded (a) significantly higher proficiency estimates than the CAT, even when item feedback was not given, and (b) significantly lower post-test state anxiety. Thus, the findings of Wise et al. (1992) were replicated and the mean proficiency estimate and anxiety differences between the self-adapted test and the CAT were found when item feedback was absent.

The studies described above indicate that a SAT has typically yielded higher mean examinee test performance than a CAT, and has been accompanied by lower mean post-test state anxiety. It is not clear, however, *why* higher test performance occurs with a SAT than with a CAT. The purpose of this study was to gather additional information regarding the dynamics of self-adapted testing.

A plausible explanation for the effectiveness of a SAT involves the concept of *perceived control*. There have been numerous studies in the psychological literature that have found that, in a stressful situation, if people believe that they have some control over the stress, they exhibit improved performance on cognitive tasks, lower anxiety, and increased motivation. An overview of this research is provided by Perlmuter and Monty (1977).

Assuming that the testing situation is stressful and examinees who are given an opportunity to choose item difficulty levels perceive that they have control over the stressful situation, the results found in previous SAT studies can be explained.

In the current study, three experimental conditions were compared. Examinees were either (a) administered a CAT, (b) administered a SAT, or (c) allowed to choose whether they wanted to be administered a CAT or a SAT. The third condition was included for two reasons. First, if the positive effects associated with a SAT are due to increases in examinee perceived control, then providing examinees with a choice between test types should enhance perceived control and possibly improve test performance. Second, by studying the test type choices made by examinees, useful information might be gained regarding the dynamics of self-adapted testing.

Research Questions

There were several research questions investigated in this study. First, does providing examinees with a choice between SAT and CAT affect test performance when compared with being assigned to a SAT or a CAT? Second, what variables influence examinee choice for SAT versus CAT? Third, what are the relative influences of test type and test choice on examinee anxiety?

Method

Examinees

A total of 377 students from a large midwestern university participated in this study. All students were enrolled in an introductory statistical methods course; data were collected from 11 course sections during the spring semester and summer sessions of 1992. The group of examinees consisted of 244 undergraduates and 133 graduate students. There were 250 females and

127 males in the sample. Examinees were randomly assigned to the three experimental conditions used in the study.

At the beginning of the statistics course, students are routinely tested to assess their working knowledge of the basic algebra skills that would be needed in the course. Students exhibiting low scores on this test were required to attend review sessions held early in the course.

Instruments

The primary instrument used was a computer-based algebra test administered using the MicroCAT testing software (Assessment Systems Corporation, 1988). Each examinee received 20 multiple-choice items drawn from a 91-item pool, with proficiency estimated using a maximum-likelihood method. The algebra test was administered in either a CAT or SAT format. Detailed information regarding item pool development, IRT model fit, and test instructions are provided in Wise et al. (1992).

The three experimental conditions used in this study were termed CAT, SAT, and CHOICE, respectively. In the CAT condition, examinees were administered a 20-item computerized adaptive test. Examinees in the SAT condition were administered a 20-item self-adapted test. In the CHOICE condition, examinees were asked to choose, prior to testing, whether they wished to receive a CAT or a SAT. In making this choice, each examinee was given the following instructions:

> Before you begin the test, you must choose how the item difficulty levels will be selected. You can either select the difficulty level of each item or let the computer select items that it judges to be of appropriate difficulty for you.
>
> Which would you like to do?
>
> A. Be allowed to select the difficulty levels of my own test items.
> B. Let the computer select the difficulty levels of my items.

After each examinee chose a testing format, he/she was then routed to either a SAT or CAT for the remainder of the testing session. In each testing format, item feedback was provided and no time limit was imposed during testing.

There were two additional instruments used in this study. The Revised Mathematics Anxiety Rating Scale (RMARS; Plake & Parker, 1982) was used to measure examinee anxiety toward mathematics. In addition, the State Anxiety Scale of the State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970) was used as a measure of situation-specific anxiety both before and after the testing session.

Procedure

The testing was completed at the beginning of the course—during the first week of the spring semester and during the first two days of the five-week summer sessions. During the first class session, students (a) were informed that the algebra test scores would be used to identify students needing review, (b) signed up for a time to be administered the algebra test, and (c) completed the RMARS.

Students were tested in groups ranging in size from 1 to 12 in a quiet room containing 12 IBM PS/2 Model 55SX microcomputers. It was prearranged that each of the test types would be administered on specific computers. On three computers, examinees were assigned to be administered the CAT; on three other computers, examinees were assigned to be administered the SAT. On the remaining six computers, examinees participated in the CHOICE condition. This oversampling of the CHOICE condition was purposeful; it yielded sufficient data to study in more detail the test type choices made by examinees and the effects of those choices. Upon arrival at the testing room, each student was directed by the test administrator to select a microcomputer. The student was assigned to a treatment condition

by his/her computer choice. This process was essentially random; the computers administering each testing format were randomly designated. Moreover, at several points during the student testing the computers were randomly redesignated.

After being seated at a microcomputer, each examinee completed a paper-and-pencil version of the State Anxiety Scale. Next, the student completed the computer-based algebra test. Pencils and scratch paper were provided and calculators were not allowed. After completing the algebra test, the State Anxiety Scale was again administered. Finally, the examinee was informed, based on his/her proficiency estimate, whether or not a review session on algebra skills would be required.

Data Analysis

The first part of the data analysis concerned comparisons among the treatment groups. Two dependent variables were used: estimated proficiency and post-test state anxiety. The primary independent variable was test type (CAT, SAT, CHOICE). In addition, math anxiety (as measured by the RMARS) was used as a blocking variable. The distribution of examinee math anxiety was divided into three groups (low, moderate, high) that contained roughly equal numbers of examinees. Hence, any reference in this study to "low" or "high" math anxiety levels should be interpreted relative to the examinees in this study, and not in an absolute sense.

The data for estimated proficiency and post-test state anxiety were each analyzed using a two-factor analysis of variance. The effects of test type were analyzed using two planned contrasts. The first contrast compared the CAT and SAT conditions; this contrast represented a replication of the Wise et al. (1992) analysis. The second contrast compared the two assigned conditions

(CAT and SAT) with the CHOICE condition. Interactions between the contrasts and math anxiety were tested as partial interactions (Keppel, 1991).

The second part of the data analysis focused on the examinees in the CHOICE condition. The relationship between math anxiety level and choice of test type was studied. A chi-square test of independence was used to test the significance of this relationship.

In the third part of the data analysis, differences between pre-test and post-test state anxiety were studied. A two-factor analysis of variance was performed with factors defined by (a) whether examinees were administered the CAT or the SAT and (b) whether examinees were assigned to or chose their test type. A .05 level of significance was used in all analyses.

Results

Treatment Group Comparisons

Table 1 contains means and standard deviations for examinee proficiency, broken down by test type and math anxiety level.

Table 1

Descriptive Statistics for Examinee Proficiency, By Test Type and Math Anxiety Level

Math Anxiety Level	CAT			SAT			CHOICE		
	Mean	SD	n	Mean	SD	n	Mean	SD	n
Low	0.60	0.83	35	0.86	0.88	29	0.72	0.74	59
Moderate	0.55	0.79	27	0.58	1.14	28	0.29	0.89	69
High	-0.63	1.09	33	-0.69	0.76	36	-0.24	0.87	60
All Examinees	0.16	1.08	95	0.17	1.15	93	0.26	0.92	188

The results of the ANOVA for examinee proficiency are shown in Table 2. Neither of the planned contrasts were significant as main effects. There was, however, a significant interaction between contrast 2 and math anxiety level. A graph of the interaction between contrast 2 and math anxiety level is shown in Figure 1. Tests of simple effects, also shown in Table 2, revealed that contrast 2 was significant only for the high anxiety examinees. For these examinees, mean proficiency was higher for examinees in the CHOICE condition than for those in the assigned conditions.

Table 2

Analysis of Variance for Examinee Proficiency

Source	SS	df	MS	F	F-Prob.
Contrast 1	.27	1	0.27	.34	.558
Contrast 2	.22	1	.22	.29	.593
Contrast 2 at Low Anxiety	.00	1	.00	.00	.954
Contrast 2 at Moderate Anxiety	2.26	1	2.26	2.91	.089
Contrast 2 at High Anxiety	5.87	1	5.87	7.55	.006
Anxiety Level	101.72	2	50.86	65.47	<.001
Contrast 1 x Anxiety	.91	2	.45	.58	.558
Contrast 2 x Anxiety	7.85	2	3.92	5.05	.007
Error	285.10	367	.78		

Note: Contrast 1 compared CAT with SAT; Contrast 2 compared CAT and SAT with CHOICE.
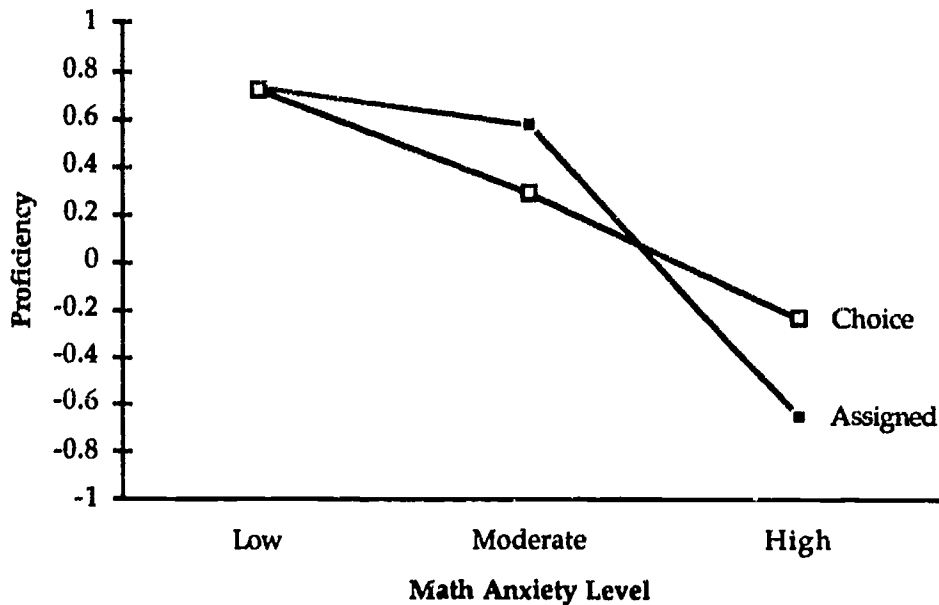
Figure 1: Proficiency Interaction Between Math Anxiety Level and the Choice vs. Assigned Groups in Contrast 2

The means and standard deviations for post-test state anxiety are shown in Table 3. Across all math anxiety levels, mean anxiety was lowest for the CHOICE condition, followed by SAT and then CAT. These differences, however, were not statistically significant as indicated by the ANOVA results shown in Table 4. Neither of the planned contrasts were significant, nor were their interactions with math anxiety level.

<u>Analysis of Examinee Choice</u>

The second part of the data analysis focused on the 188 examinees in the CHOICE condition. Figure 2 shows the numbers of examinees choosing CAT and SAT at each level of math anxiety. Examinees low in math anxiety showed a strong preference for CAT. As anxiety level increased, however, there was a corresponding increase in preference for SAT; the majority of the examinees reporting high math anxiety chose SAT.

Table 3

Descriptive Statistics for Post-Test State Anxiety, By Test Type and Math Anxiety Level

| Math Anxiety Level | Test Type | | | | | | | | |
| | CAT | | | SAT | | | CHOICE | | |
	Mean	SD	n	Mean	SD	n	Mean	SD	n
Low	32.59	9.99	34	32.31	9.93	29	32.95	7.39	59
Moderate	41.11	10.08	27	40.89	10.75	27	39.18	11.25	68
High	50.64	10.91	33	46.72	13.15	36	44.88	11.11	60
All Examinees	41.37	12.79	94	40.47	12.91	92	39.04	11.16	187

Table 4

Analysis of Variance for Examinee Post-Test State Anxiety

Source	SS	df	MS	F	F-Prob.
Contrast 1	99.39	1	99.39	.89	.345
Contrast 2	269.28	1	269.28	2.42	.121
Anxiety Level	12667.35	2	6333.67	56.93	<.001
Contrast 1 x Anxiety	145.18	2	72.59	.65	.521
Contrast 2 x Anxiety	288.13	2	144.07	1.30	.275
Error	40493.55	364	111.25		

Note: Contrast 1 compared CAT with SAT; Contrast 2 compared CAT and SAT with CHOICE.
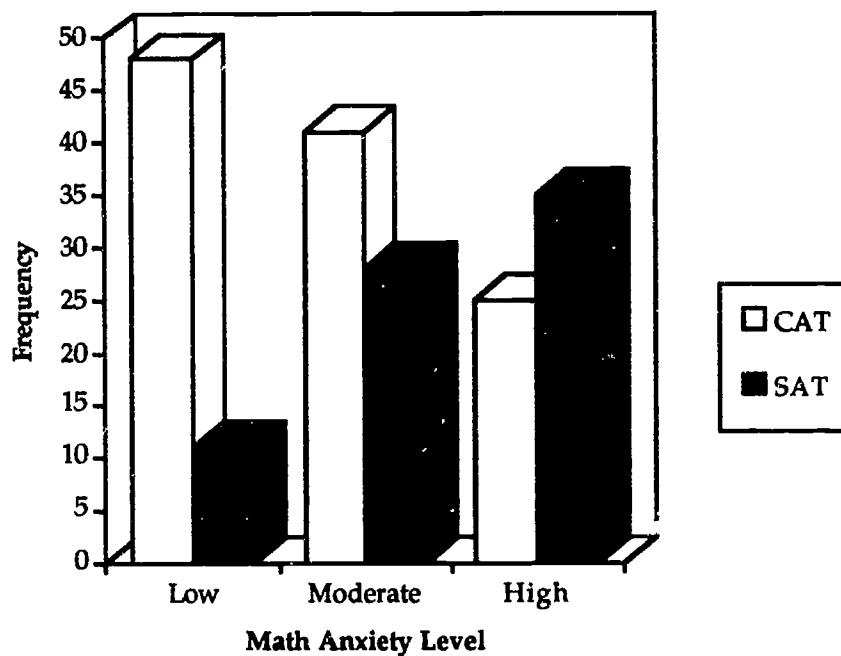
Figure 2: Frequency of Examinee Choice of Each Test Type, By Level of Math Anxiety

The chi-square test of independence found a highly significant relationship between test choice and math anxiety ($\chi^2 = 19.701$, df = 2, p < .0001).

Anxiety Difference Scores

Difference scores between pre-test and post-test state anxiety were formed. Table 5 shows the means of standard deviations of these scores for the four groups defined by test type and whether the test was assigned or chosen. When the SAT was chosen, mean anxiety showed a slight decrease from pre-test to post-test. In the other three groups, mean anxiety increased. The results of the ANOVA for these data are given in Table 6. Only the main effect for test type was found significant. Examinees receiving the CAT exhibited a larger increase in state anxiety than those receiving the SAT.

Table 5

Descriptive Statistics for Difference Between Pre-Test and
Post-Test State Anxiety, by Test Type and Choice Condition

Group	Mean	SD	n
CAT, Assigned	-3.59	10.26	94
CAT, Chosen	-2.90	8.90	114
SAT, Assigned	-0.90	8.61	93
SAT, Chosen	0.51	9.52	73

Note: A negative mean indicates an increase in reported anxiety
during testing; a positive mean indicates a decrease in reported anxiety.

Table 6

Analysis of Variance for Difference Between Pre-Test and Post-Test State
Anxiety

Source	SS	df	MS	F	F-Prob.
Test Type	846.19	1	846.19	9.76	.002
Choice Condition	99.75	1	99.75	1.15	.284
Test x Choice	12.10	1	12.10	.14	.701
Error	32081.13	370	86.71		

Discussion

It was found, for examinees reporting high math anxiety, that
providing a choice between CAT and SAT led to significantly higher mean
proficiency estimates. This finding represents support for the hypothesis that
examinees can more effectively cope with a stressful situation if they feel they
have some control over the source of stress. It also suggests that highly

anxious examinees would benefit the most from increased control over the testing situation. Interestingly, the significantly higher test performance for highly math anxious examinees was not paired with significantly lower post-test state anxiety, as was found in the Wise et al. (1992) and the Roos et al. (1992) studies.

The expected differences in estimated proficiency and post-test state anxiety between the CAT and the SAT conditions were not found. These results are curious, because this study's CAT-SAT comparison represents a direct replication of the Wise et al. (1992) study. The testing procedures, item pool, and examinee population were all the same in the two studies. Moreover, the Roos et al. (1992) did replicate the Wise et al. study under the same testing conditions. Although it is tempting to interpret the nonsignificant CAT-SAT differences found in the current study as a Type II error, it should be kept in mind that relatively few CAT-SAT comparison studies have been conducted thus far. As additional studies are completed, interpretation of the current study's results should become more clear.

A strong relationship was found between examinee test type choice and math anxiety level. It appears that the SAT was most attractive to the highly math anxious examinees. For the less math anxious examinees, the CAT was the more popular choice. It is interesting to note that many examinees, when given the opportunity to gain greater control over the testing situation by being allowed to select their item difficulty levels, chose not to have that control. A possible explanation for these findings is that examinees are not motivated to accept control when they do not perceive the testing situation as sufficiently stressful. In the current study, the consequences for poor test performance (attending an algebra review session) were not very severe; one might speculate that a higher-stakes testing situation would be perceived as

highly stressful by a larger proportion of the examinees. In this case, the SAT should become attractive to more examinees. More research is needed on the relationship between examinee perception of stress and preference for a SAT.

The analysis of the state anxiety difference scores indicated a joint effect of control of test type and control of item difficulty level; the highest mean difference score was found when both forms of control were provided. Only the main effect for test type was significant, however, suggesting that control over item difficulty was more important than having the opportunity to choose test type. This finding may be related to the number of choices available to an examinee. The choice of test type could be made only once, while the choice of item difficulty level could be made 20 times. Examinee feelings of control may increase as more choice opportunities are provided.

Conclusions

The results of the current study support the hypothesis that increasing an examinee's perception of control over a testing situation can have positive effects on test performance. This control hypothesis would readily explain the results of previous studies that have shown examinees administered a SAT perform higher than examinees administered a CAT. This study also found evidence that higher anxiety examinees have a greater preference for the control provided by a SAT.

It is becoming increasingly clear that the use of computers in testing provides opportunities for more effective measurement. While it has been well established that a CAT can provide more efficient measurement, a SAT holds promise for providing more *valid* measurement. If providing examinees control over their item difficulty levels reduces the influence of test anxiety on estimated proficiency, then the resulting scores should be more valid measures of proficiency. Rocklin and O'Donnell (1991) and Wise

(1992) provide evidence that the influence of anxiety is reduced when a SAT is used. Evidence for the increased validity of SAT-based proficiency estimates, however, has not yet been found. This issue should be of primary concern in future investigations of self-adapted testing.

References

Assessment Systems Corporation. (1988). *User's manual for the MicroCAT testing system, version 3.* St. Paul, MN: Author.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Engelwood Cliffs, NJ: Prentice-Hall.

Perlmuter, L. C., & Monty, R. A. (1977). The importance of perceived control: Fact or fantasy? *American Scientist, 65,* 759-765.

Plake, B. S., & Parker, C. S. (1982). The development and validation of a revised version of the mathematics anxiety rating scale. *Educational and Psychological Measurement, 42,* 551-557.

Rocklin, T., & O'Donnell, A. M. (1987). Self-adapted testing: A performance-improving variant of computerized adaptive testing. *Journal of Educational Psychology, 79,* 315-319.

Rocklin, T., & O'Donnell, A. M. (1991, April). *An empirical comparison of self adapted and maximum information item selection.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Roos, L. L., Plake, B. S., & Wise, S. L. (1992, April). *The effects of feedback in computerized adaptive and self-adapted tests.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Vispoel, W. P., & Coffman, D. D. (in press). Computerized adaptive and self-adapted music listening tests: psychometric features and motivational benefits. *Applied Measurement in Education*.

Wise, S. L. (1992, June). *Self-adapted testing: Promise and implications*. Paper presented at the European Conference on Educational Research, University of Twente, The Netherlands.

Wise, S. L., Plake, B. S., Johnson, P. L., & Roos, L. L. (1992). A comparison of self-adapted and computerized adaptive tests. *Journal of Educational Measurement*, 29(4), 329-339.