INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

RICE UNIVERSITY

**Computer Adaptive Testing: The Impact of Test Characteristics
on Perceived Performance and Test Takers' Reactions**

by

Scott Tonidandel

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Miguel A. Quinones, Chair
Associate Professor of Psychology

Richard P. Bagozzi
Professor of Management

Robert L. Dipboye
Professor of Psychology

Michelle R. Hebl
Assistant Professor of Psychology

HOUSTON, TEXAS

MAY, 2001

UMI Number: 3021197

UMI Microform 3021197

Copyright 2001 by Bell & Howell Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

MAY, 2001

ABSTRACT

Computer Adaptive Testing: The Impact of Test Characteristics on
Perceived Performance and Test Takers' Reactions

by

Scott Tonidandel

This study examined the relationship between characteristics of adaptive testing and test takers' subsequent reactions to the test. Participants took a computer adaptive test in which two features, the difficulty of the initial item and the difficulty of subsequent items, were manipulated. These two features of adaptive testing determined the number of items answered correctly by examinees and their subsequent reactions to the test. The data show that the relationship between test characteristics and reactions was fully mediated by perceived performance on the test. In addition, the impact of feedback on reactions to adaptive testing was also evaluated. In general, feedback that was consistent with perceptions of performance had a positive impact on reactions to the test. Implications for adaptive test design concerning maximizing test takers' reactions are discussed.

Table of Contents

INTRODUCTION	1
Computer Adaptive Testing	2
Test Characteristics	6
Number of Items Answered Correctly	11
Perceptions of Performance	12
Test Takers' Reactions	13
Post Feedback Reactions	22
METHOD	25
Development of the Adaptive Test	25
Subjects	29
Design	30
Measures	30
Experimental Conditions	33
Procedure	34
RESULTS	35
Number of Items Correct	35
Perceived Performance	39
Motivation	39
Test Anxiety	40
Satisfaction	40

Attributions	41
Self-efficacy	44
Fairness	46
Feedback Acceptance	46
Post Feedback Reactions	47
Experience with Adaptive Tests	56
DISCUSSION	56
REFERENCES	65
APPENDIX 1: FACTOR STRUCTURE AND LOADINGS OF ITEMS	71

List of Tables and Figures

Table 1: Descriptive statistics and correlation coefficients for the study variables.

Table 2: Mean Number of Questions Answered Correctly by Participants in Each Condition.

Table 3: Tests of the Interaction of Self-efficacy and Perceived Performance on Attributions of Performance.

Table 4: Tests of the interaction of actual performance X perceived performance X feedback acceptance on post feedback fairness perceptions.

Figure 1: Proposed model illustrating the relationship between characteristics of an adaptive test, perceptions of performance, and test takers' reactions.

Figure 2: Interactions of Perceived Performance X Self-efficacy on Attributions of Performance to the Task.

Figure 3: Interaction of perceived performance X actual performance on feedback acceptance.

Figure 4: Interaction of actual performance X feedback acceptance on post feedback satisfaction.

Figure 5: Interaction of actual performance X perceived performance X feedback acceptance on post feedback fairness perceptions.

Introduction

Continuing improvements in technology have lead to the proliferation of extremely powerful computers that are available at relatively low prices. Due to the increased affordability and computational ability of modern computers, computerized assessment has gained a more prominent role in testing. In particular, computer adaptive testing, a method of testing that tailors the administration of test questions to the ability of each examinee, is being adopted for both selection and assessment purposes. For example, in the public sector, the Educational Testing Service now offers only a computer adaptive version of the Graduate Records Examination (GRE), and the United States military currently assesses at least two-thirds of their recruits using a computer adaptive version of the Armed Service Vocational Aptitude Battery (ASVAB; Segall & Moreno, 1999). In private industry, State Farm Insurance has developed an adaptive test for the selection of entry-level computer programmers (Zickar, Overton, Taylor, & Harms, 1999). This newfound reliance on computer adaptive testing necessitates an examination into the efficacy of using adaptive testing as a selection device. Although, the psychometric properties of adaptive testing have been well documented, the psychological impact this new mode of assessment has on the examinee is unknown. The purpose of this study is to examine how specific features of adaptive tests are related to test taker reactions to the test.

Computer Adaptive Testing

The rise in popularity of adaptive testing most likely is attributable to advantages such as increased test security, instantaneous score reporting, enhanced efficiency, and superior measurement precision. Test security is improved by using an adaptive test because the process of tailoring the presentation of items to individual examinees ensures that test takers will not receive the same items (Zickar et al., 1999). Since it is unlikely that any two examinees will receive the same set of items, copying answers from another test taker is pointless. Similarly, attempts to cheat by learning what test questions were presented to a previous examinee also will be unsuccessful. Because potential examinees are unable to discern from previous examinees the test questions that will be presented to them, adaptive tests can be administered to examinees at different times without fear of the test being compromised. The increased security afforded by adaptive testing provides a tremendous advantage in terms of flexibility of administration. In order to achieve a comparable level of security on a paper-and-pencil test, a parallel test for every examinee would be required.

Adaptive tests also hold much promise for reducing testing time. Mislevy and Bock (1989) suggest that adaptive testing enables testing time to be reduced to one-half or one-third of that required for a conventional test of the same precision. This reduced testing time is accomplished by administering items that make the largest contributions to estimating examinee ability to an examinee and eliminating those items that provide little or no information about examinee ability. Because adaptive tests are able to achieve

comparable ability estimates with fewer items, the tests are more efficient than traditional paper-and-pencil tests. Aside from enhancing the efficiency of a test, the capability of adaptive testing to administer the most informative items also provides superior measurement precision. The standard error of measurement of a test is inversely proportional to the amount of information about differences between examinees that a test provides. Because the items that comprise an adaptive test provide more information about differences between examinees on average, the standard error of measurement of an adaptive test is smaller, and thus, the estimate of an examinee's ability is more precise. The superior precision and efficiency afforded by adaptive testing has been well documented in numerous studies (see Weiss, 1982, for a review).

In addition to enhanced efficiency and precision, adaptive tests are also considered to provide better measurement quality compared to conventional tests, because adaptive tests permit the customization of measurement precision to a targeted level of ability (Zickar et al., 1999). On a conventional test, measurement precision is dependent on an examinee's ability level. For example, the test may be very precise for individuals of moderate ability, but very imprecise for individuals of very high or low ability. In contrast, adaptive tests can be structured such that the measurement precision of the test is the same for all examinees.

Other benefits of adaptive testing, such as instantaneous scoring and increased standardization, result from being administered via computer (Wainer, 1990). The computer can calculate a test taker's score and provide feedback immediately after a test

has terminated. Additionally, the computer can present all the instructions insuring that the procedures used to administer the test are identical across testing sessions. These aspects of computerized administration produce the added benefit of reducing the manpower needed to proctor and score the tests.

Numerous studies have evaluated the comparability of individuals' scores on paper-and-pencil and adaptive tests (Legg & Buhr, 1992; Overton, Harms, Taylor, & Zickar, 1997; Vispoel, Rocklin, & Wang, 1994). Similarly, a multitude of research has explored the effects of different characteristics of adaptive testing on examinee ability estimates (see Hambleton, Zaal, & Pieters, 1991 for a review). Despite the abundance of research on adaptive testing, nearly all of it is focused solely on evaluating the psychometric properties of adaptive tests. Essentially, this research is only concerned with maximizing the precision and efficiency through which adaptive tests can estimate an examinee's ability.

Because of the psychometric emphasis of the research, few studies have endeavored to examine the psychological impact of such a divergent mode of assessment. Nevertheless, the psychological impact of adaptive testing may be vital to consider independent of the psychometric characteristics of the test. Researchers have recognized that selection procedures, even those with comparable validities, may have differential effects on applicant reactions (Smither, Millsap, Stoffey, Reilly, & Pearlman, 1996). In addition, the importance of reactions to selection procedures has been well documented in the personnel selection literature, and are related to a number of important

organizational outcomes such as satisfaction with the organization, satisfaction with the job, intentions to accept a job offer, willingness to continue with the selection process, motivation to pursue the job, willingness to recommend the company to others, turnover intentions, and performance on the job (Gilliland, 1994; Konovsky & Cropanzano, 1991; Linden & Parsons, 1986; Macan, Avedon, Paese, & Smith, 1994; Rynes & Connerly, 1993; Rynes, Heneman, & Schwab, 1980; Smither Reilly, Millsap, Pearlman, & Stoffey, 1993). In addition, reactions to selection procedures may affect the validity and utility of a selection procedure due to the loss of qualified applicants and may also be related to the likelihood of litigation (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997).

The preceding research suggests that psychological reactions to selection procedures are related to important organizational outcomes. For these reasons, the psychological impact of adaptive testing should be of paramount concern to test developers wishing to use adaptive test for selection or assessment purposes. The goal of the present study is to examine how characteristics unique to adaptive testing may potentially influence examinees' reactions. Two features in particular that may affect test taker reactions are the difficulty of the initial item and the difficulty of subsequent items presented to an examinee. The present study assesses the impact of these two features on six reactions that may be important to consider when evaluating an adaptive test: motivation during the test, anxiety during the test, satisfaction with the test, attributions of performance, self-efficacy, and perceived fairness of the test. Figure 1 contains the proposed framework depicting hypothesized linkages between test characteristics and

psychological reactions. As illustrated in Figure 1, characteristics of an adaptive test are hypothesized to determine the number of questions answered correctly. The number of questions answered correctly is predicted to subsequently affect test takers' reactions toward the test by influencing examinees' subjective perceptions of performance on the test. The remainder of the paper will address the various relationships implied by the proposed framework.

Test Characteristics

Adaptive tests differ from paper-and-pencil tests in numerous ways. On paper-and-pencil tests, the same number of items are administered to all examinees. In contrast, on adaptive tests, the number of items administered depends on the criteria used to terminate the test (Thissen & Mislevy, 1990). Some adaptive tests require that a certain number of items be answered by all examinees. On an adaptive test structured in this manner, similar to a conventional test, all examinees receive the same number of questions. An alternative stopping procedure requires that a certain level of measurement precision be attained before the test is terminated. Because the number of items required to reach a certain level of measurement precision will vary across individuals, the implication of this stopping procedure is that different test takers will likely receive a different number of items.

Figure 1

Proposed Framework Illustrating the Relationship Between Characteristics of an Adaptive Test, Perceptions of Performance, and Test Takers' Reactions

Adaptive tests also differ from conventional tests in terms of an examinee's ability to navigate within the test (Wise, 1997). On an adaptive test, test takers are typically required to answer every question in the order they are presented. They are not allowed to skip questions and return to them later, change answers on previous questions, nor are they able to review their answers to previous questions. Each of these options usually is available to examinees on conventional tests.

Paper-and-pencil tests and adaptive tests also differ in the procedure used to calculate an examinee's score (Stocking, 1996). Traditional tests are typically based on classical test theory, and an individual's score is computed by summing the number of questions that are answered correctly. Adaptive tests, on the other hand, are based on item response theory. According to item response theory an individual's score is not based solely on the number of items answered correctly but rather, on which items were answered correctly. In general, individuals are given more credit for answering difficult questions correctly than for answering easier questions. Items also are weighted in terms of the amount of information they provide (Stocking, 1996).

Other differences between conventional and adaptive tests arise from decisions of how to start an adaptive test and which item to administer next in the test. For example, adaptive test designers can choose to begin a test with an easy or a difficult item. Similarly, the test developer can manipulate the difficulty of subsequent items. The present study is concerned specifically with how these two aspects of adaptive testing, the

difficulty of the initial item and the difficulty of subsequent items that are selected for administration, may affect test takers' reactions.

Difficulty of the initial item. Traditional paper-and-pencil tests often start with easy items as a warm-up for examinees (Bergstrom & Lunz, 1999). In contrast, adaptive tests frequently begin by administering an item of moderate difficulty. When adaptive tests are designed for licensure or certification purposes, the passing score usually determines the difficulty of the initial item. If the criterion for passing the test is high, the initial item presented to examinees may be exceedingly difficult. Research by Lunz, Bergstrom, and Gershon (1994) indicates that performance on an adaptive test is unaffected by the difficulty of the initial item. In other words, regardless of whether an adaptive test begins with an item of low, moderate, or high difficulty, examinees' ability estimates appear to be identical.

Although performance seems to be unaffected by the difficulty of the initial items, the difficulty of the starting item may potentially impact examinees' reactions. By beginning with easy items, paper-and-pencil tests allow most test takers to begin the test with success experiences, and some test developers believe that this initial success influences reactions (Mills & Stocking, 1996; Mills, 1999). However, the mechanism through which the difficulty of the initial item may affect reactions is unknown. In addition, the potential benefits of early success experiences on test takers' reactions have not yet been confirmed in the literature. Thus, one objective of the present study is to

evaluate the impact of initial item difficulty on test takers' reactions toward an adaptive test.

Difficulty of subsequent items. A common procedure for selecting which item to administer next in an adaptive test is maximum information (Thissen & Mislevy, 1990). Under conditions when guessing is minimal, an item provides maximum information when ability is equal to the item's discrimination parameter (Hambleton, Swaminathan, & Rogers, 1991). The item discrimination parameter is equal to the point on the item characteristics curve (ICC) where the probability of a correct response is fifty-percent. Since the maximum information item selection procedure selects items for an individual so that he/she will have a fifty-percent chance of getting that item correct, over the course of an entire adaptive test, examinees will answer approximately fifty-percent of the items correctly. In contrast, paper-and-pencil tests, specifically those exams used for licensing purposes, usually require that examinees answer approximately seventy-percent of the test items correctly in order to pass. Hence, the proportion of items answered correctly on an adaptive test is frequently far lower than examinees are used to experiencing with a conventional test (Wise, 1997; Lunz et al., 1994). Consequently, adaptive tests with a fifty-percent probability of answering items correctly may be perceived as an unduly hard test by conventional test standards.

Sutton (1997) notes that high performing examinees are accustomed to answering a high proportion of items correctly on traditional tests. These examinees may not be familiar with the more difficult adaptive tests, and they may find such tests to be

frustrating (Bergstrom & Lunz, 1999). Previous research has demonstrated that the increased difficulty of adaptive tests has little impact on test takers' actual performance (Bergstrom, Lunz, & Gershon, 1992; Lunz et al., 1994). However, the effect that this change in difficulty level has on motivation, attitudes, and expectancies is not clear. Bergstrom et al. (1992; Bergstrom & Lunz, 1999) suggest that altering the probability of a correct response may reduce potentially adverse psychological reactions that may result from an extremely difficult test. Despite this supposition, no empirical evaluation has been conducted to assess the extent to which a difficult adaptive test elicits adverse reactions. Additionally, the purported benefits of making the test easier by modifying the item selection procedures are also unsubstantiated. Thus, a second aim of this study is to systematically examine how the difficulty of the items administered in an adaptive test affects test takers' reactions toward the test.

Number of Items Answered Correctly

Although adaptive tests differ from paper-and-pencil tests along a number of dimensions, the present study is concerned solely with the difficulty of the initial item and the difficulty of subsequent items presented to examinees. The research cited above illustrates that these two features of adaptive testing can be manipulated without compromising the psychometric properties of the test. Nevertheless, the difficulty of the initial item and the difficulty of subsequent items may be particularly important in shaping test takers' reactions because these two features play a large role in determining the number of questions answered correctly by a test taker. For example, examinees who

are administered easy items (either initially or subsequently) will get more items correct than those who received more difficult questions.

H1: Characteristics of an adaptive test will determine the number of questions answered correctly. Specifically, the difficulty of the initial item and the difficulty of subsequent items will predict the number of questions answered correctly independent of test taker ability.

Perceptions of Performance

On a traditional paper-and-pencil test, an individual's score is based on the number of questions answered correctly. In contrast, on an adaptive test, an individual's score is based on which items are answered correctly rather than the number of items answered correctly. In other words, an individual who answers difficult questions will score higher than an individual who answers easier questions, even if they get the same number of questions correct. This new method of scoring tests may have a profound impact on examinees' subjective perceptions of performance. If examinees base estimates of their performance on the number of questions answered correctly, they should be considerably less accurate in gauging their performance on an adaptive test compared to a paper-and-pencil test. The literature tends to support this assertion. Macan et al. (1994) found the correlation between examinees' perceived performance on paper-and-pencil cognitive ability test and their actual score on the test to be $r = .40$. In contrast, for adaptive tests, the correlation was only $r = .069$ (Powell, 1994). Lending further support, Betz and Weiss (1976) examined the relationship between ability and perceived test

difficulty across two groups of college students taking either an adaptive or non-adaptive computer administered test. They found a significant difference in the correlation between the two groups. Because, adaptive tests administer items of appropriate difficulty to examinees of different ability levels, the adaptive test was perceived as equally difficult by both high and low ability participants. In contrast, on the non-adaptive test where the same items were administered to all examinees, low ability examinees reported that the test was more difficult than did high ability examinees.

The aforementioned research demonstrates that characteristics of a test may be potentially important determinants of test takers' subjective perceptions of performance. If examinees base performance estimates on the perceived number of questions answered correctly, features of adaptive tests that decrease the number of questions answered correctly will influence perceptions of performance. The following hypothesis is offered: H2: The number of questions answered correctly will positively predict perceived performance independent of test taker ability.

Test Takers' Reactions

Test takers' perceptions of their performance may be a potential mediator of the impact of test characteristics on subsequent reactions to the test. Powell (1994) compared the effect of three different types of item selection procedures on reactions toward the test. The procedures evaluated in that study were a conventional computer adaptive test in which the difficulty of items selected were based on ability estimates, a self-adaptive test in which examinees could choose the difficulty of the next item presented, and finally, a

random procedure in which the next item was selected randomly from the test bank. The results indicated that these different methods of item selection did not affect test performance. However, reactions to these three different modes of testing were related to perceived, but not actual, levels of performance. The preceding study implies a potential mediating influence of perceptions of performance on the relationship between test characteristics and test takers' subsequent reactions.

Motivation. Betz and Wiess (1976) performed the most extensive evaluation of the effect of adaptive testing on examinee motivation. In this study, two groups of high and low ability college students took either an adaptive or non-adaptive test, and in-test motivation across these four conditions was compared. For the high ability group, motivation was lower for the adaptive test when compared to the conventional test, although the difference was not statistically significant. For the low ability students, motivation was significantly higher for those taking the adaptive test than for those taking the non-adaptive test. These findings can be interpreted in light of students' perceptions of test difficulty. High ability students reported that the conventional test was easier compared to the adaptive test, whereas low ability students gave lower difficulty ratings for the adaptive test compared to the non-adaptive test. Together, these results suggest that self-reported motivation may be higher for tests that are perceived as easier than tests that are viewed as more difficult. If people believe that they perform better on easier tests, then features of an adaptive test that influence perceptions of performance will likely also influence motivation.

H3: Perceived performance on the test will mediate the relationship between test characteristics and test-taking motivation.

Anxiety. Anxiety in a testing situation may be a function of both trait and state components, and thus, anxiety might be altered to some extent by the testing situation (Hembree, 1988). Consequently, examining features of adaptive tests that may reduce anxiety is important. Test anxiety has been found to be negatively correlated to test performance (see Hembree, 1988, and Seipp, 1991, for meta-analyses). As a result, variables that elicit feelings of success or failure, such as perceived performance, will likely be correlated with anxiety as well. In support, Powell (1994) examined the relationship between perceived performance and anxiety on an adaptive test and found a negative relationship that approached conventional levels of significance ($p = .053$).

H4: Perceived performance will mediate the relationship between test characteristics and test anxiety.

Satisfaction. Applicants' satisfaction with a selection procedure has been linked to a number of outcomes such as organizational attractiveness, willingness to recommend the job to others, and intentions to accept a job offer (Linden & Parsons, 1986; Smither et al., 1993). The literature on attitudes toward adaptive testing suggests that examinees are generally satisfied with adaptive testing procedures. For example, a study by Vispoel et al. (1994) asked examinees to evaluate specific attributes of adaptive testing. They found that participants liked the shorter length of adaptive tests and the ability to receive feedback. However, participants disliked the inability to review or skip items on adaptive

tests. Moe and Johnson (1986) examined the attitudes of 315 high school students toward adaptive testing. These students had favorable impressions of adaptive testing and reported a preference for adaptive tests over paper-and-pencil tests, mainly because they liked not having to fill out traditional bubble forms and the shorter length of the adaptive test. Similarly, other researchers have also found that adaptive test takers generally satisfied with the testing experience (Schmidt, Urry, & Gugel, 1978; Legg & Buhr, 1992). However, these studies used primarily survey methodology and failed to examine test takers' satisfaction across tests with different characteristics. Thus, it remains unclear whether the favorable reactions to adaptive testing are more or less positive than would have been found with a different type of test.

Satisfaction with a selection test appears to be related to performance. As expected, Lonsbury, Bobrow, and Jansen (1989) found that individuals who failed a pre-employment test had less favorable attitudes toward employment testing than individuals who passed. Similarly, perceived performance may be closely related to satisfaction in an adaptive testing context. Powell (1994), comparing a conventional adaptive test, a self-adaptive test, and a non-adaptive test, found that perceived, but not actual, performance on these three different computer administered exams was positively related to preferences toward these tests. These results suggest that perceptions of performance might be a more important determinant of satisfaction with an adaptive test than actual performance.

H5: Perceived performance will mediate the relationship between test characteristics and satisfaction.

Attributions of performance. After performing a task, individuals often search for a cause of their performance. The causes of one's performance are ascribed typically to factors internal to the person, such as ability or effort, or to factors external to the person, such as luck or task difficulty (Weiner, 1971). Weiner (1985) presents a theory of attributions in which the type of attribution made can have both psychological and behavioral outcomes. Typically, the driving force behind attributions of performance is an individual's success or failure at a task (Weiner, 1985). In addition, expectations of performance also play an important role in determining attributions. For example, Feather and Simon (1972) found that expectancies interact with performance outcomes to determine the types of attributions that were made. When participants performed at a level that was not consistent with their expectations, they were more likely to attribute performance to external factors (luck) than when performance was consistent with expectations. Similarly, McHahan (1973) found that when levels of performance and expectancies were alike participants were more likely to make attributions to ability, while attributions to luck were more likely when performance and expectancies differed.

Self-efficacy theory (Bandura, 1997, 1982, 1986) provides a useful mechanism to explain the relationship between performance, expectancies, and attributions. Self-efficacy has been defined as a person's assessment of his/her capacity to perform on a certain task (Bandura, 1977). In a review of the self-efficacy literature, Gist and Mitchell

(1992), conclude that under conditions of failure people with high self-efficacy attribute their failure to external factors, such as bad luck, whereas people with low self-efficacy attribute failure to internal factors, such as a lack of ability. In support of this assertion, Quiñones (1995), examining the relationship between performance and attributions and found that self-efficacy was a significant moderator of the relationship. Participants high in self-efficacy were more likely to attribute performance to internal factors when successful and external factors when unsuccessful, while the reverse was true for participants low in self-efficacy. Similarly, Silver, Mitchell, and Gist (1995) found that self-efficacy moderated the relationship between self-efficacy and attributions. Although the aforementioned studies examined moderating effects of self-efficacy on the relationship between attributions and actual performance, the same relationship is anticipated for perceived performance.

H6: Perceived performance will mediate the relationship between test characteristics and attributions. In addition, the relationship between perceived performance and attributions will be moderated by self-efficacy. Specifically, the relationship between perceived performance and attributions to internal factors will be greater for individuals high in self-efficacy than for individuals low in self-efficacy. The opposite pattern of results is anticipated for attributions to external factors.

Aside from examining antecedents of attributions, the consequences of making certain types of attributions are also worth investigating. One psychological consequence of attributions may be individuals' expectations of future success. In particular, Wiener

(1985) notes that attributions to stable causes, such as ability or skill, are more likely to influence future expectations than attributions to variable causes, such as luck.

Additionally, the impact of attributions on subsequent self-efficacy may also depend on performance. Bandura (1977) described four sources of information that may influence an individual's self-efficacy: performance accomplishments, vicarious learning, verbal persuasion, and emotional arousal. According to Bandura, the most effective of these may be performance accomplishment (past performance). Bandura (1997) hypothesize that past performance might interact with attributions to affect post task self-efficacy. He surmised that successful performance would have a greater impact on self-efficacy when the success resulted from internal causes than external causes. Similarly, stable causes of performance were seen as more likely to influence subsequent self-efficacy. Silver, Mitchell, and Gist (1995) provided empirical support for this hypothesis when they found that attributions of performance moderated the relationship between past performance and post task self-efficacy.

H7: Perceived performance will mediate the relationship between test characteristics and posttest self-efficacy. In addition, the relationship between perceived performance and posttest self-efficacy will be moderated by attributions. Specifically, the relationship between perceived performance and self-efficacy will be greater when attributions are made to internal stable causes of performance than when attributions are made to external variable causes of performance.

Perceived fairness. The fairness literature typically makes a distinction between three different types of fairness: distributive, procedural, and interactional. Distributive fairness is concerned with the perceived fairness of outcomes, results, or ends that are achieved. Procedural fairness refers to the perceptions of the means used to achieve the results (i.e., the procedures). Finally, interactional justice deals with the perceived fairness of social interactions. The present study is concerned only with aspects of distributive fairness. The focus will be on whether the outcome (i.e., test score) is perceived to be fair.

The perceived fairness of a selection system has profound consequences on an applicant's subsequent attitudes and behaviors. Konovsky and Cropanzano (1991) found that the perceived fairness of an actual drug testing program predicted job satisfaction, commitment to the organization, trust in management, turnover intentions, and job performance. Similarly, Farmer, Beehr, and Love (1998) found that fairness perceptions of a police officer selection system were related to job satisfaction, organizational commitment, and subsequent job performance. In addition, other researchers have found the perceived fairness of selection procedures to be related to intentions to recommend the job to others (Smither et al., 1993; Gilliland, 1994). The literature examining the impact of adaptive testing on perceptions of fairness is somewhat mixed. Schmidt et al. (1978) examined fairness perceptions of 163 examinees taking a pilot version of an adaptive civil service exam. Respondents tended to report that the adaptive test was fair. In contrast, research by Tonidandel and Quiñones (2000) demonstrates that certain

features of adaptive testing, such as the inability to skip questions and the administration of different sets of items to test takers, are seen as unfair by potential examinees.

Gilliland (1993) presents a framework of organizational justice that describes applicant reactions to employment selection systems. Specifically, this framework attempts to identify the conditions in which applicants are likely to evaluate selections systems as unfair. According to Gilliland, applicants' judgments of fairness depend on both the outcome of the selection decision and their expectation regarding that outcome. Gilliland states "Perceptions of equity [fairness] will emerge from the matching of applicants' expectations with the outcome of the hiring decision (p 717)". In other words, to the extent that expectations about performance are consistent with perceptions of performance, perceived fairness of a selection procedure should be enhanced. In support, Quiñones (1995) examined fairness perceptions of a training placement test, and found that the relationship between perceived performance on the test and fairness perceptions was moderated by self-efficacy. For participants high in self-efficacy, the test was seen as more fair when perceived performance was high and less fair when perceived performance was low. The reverse was true for participants low in self-efficacy. To summarize, when expectations of performance were congruent with perceptions of performance, the test was viewed as more fair. Gilliland (1994) found an identical pattern of results for fairness perceptions of an actual selection system. When outcomes were consistent with expectations, applicants viewed the selection process as more fair.

H8: Perceived performance will mediate the relationship between test characteristics and fairness perceptions. In addition, self-efficacy will moderate the nature of the relationship. Specifically, the relationship between perceived performance and fairness perceptions will be positive for individuals high in self-efficacy and negative for individuals low in self-efficacy.

Post Feedback Reactions

After completion of a selection procedure, applicants are usually given some type of feedback reflecting their performance. This feedback may consist of a score on a selection test or the feedback may just reflect the final outcome of the selection process (i.e. selected or rejected for the job). The potential exists for applicant reactions to be affected by the feedback that they receive. For example, individuals who received negative feedback regarding performance on a selection procedure were more likely to report that the procedure was unfair (Gilliland, 1994; Kluger & Rothstein, 1993). In a conventional testing situation, perceptions of performance typically are closely related to actual performance on the test (Macan et al., 1994). Thus, feedback may have little additional impact on reactions. In an adaptive testing context, because the relationship between perceived and actual performance is considerably less (Powell, 1994), the potential exists for feedback that reflects actual performance to influence reactions.

In order for feedback to have an influence on reactions, the recipient must first accept the feedback. Feedback acceptance has been defined as “the recipient’s belief that the feedback is an accurate portrayal of his or her performance (Ilgen, Fisher, & Taylor,

1979). Ilgen et al. (1979) examined factors that may influence feedback acceptance. From their review of the literature, they concluded that a major characteristic of the feedback that affects acceptance is consistency. Consistency refers to the extent to which the feedback currently presented is similar to feedback previously given to an individual. In a testing context, examinees typically only receive feedback once, at test completion. Nevertheless, test takers usually have a subjective sense of how well they have performed on a test. Thus, the similarity between these subjective perceptions of performance and the feedback presented upon test completion should influence feedback acceptance. In support of this assertion, Taylor, Fisher, and Ilgen (1984) conclude that “individuals assess the accuracy of feedback they receive and the perceived accuracy of this information moderates their reactions to it (p. 97).”

H9: The relationship between feedback and feedback acceptance will be moderated by perceived performance. Specifically, individuals who receive positive feedback will be more likely to accept the feedback if they perceived that they performed well and less likely to accept the feedback if they perceived that they performed poorly. The opposite relationship is anticipated for individuals who receive negative feedback.

If the recipients do not accept the feedback, feedback will likely have no affect on subsequent reactions. However, if acceptance occurs, the potential exists for the feedback to affect test takers' reactions. Nease, Mudgett, and Quiñones (2000) found that feedback acceptance moderated the relationship between feedback and subsequent reactions. Similarly, in the present study, feedback acceptance is expected to moderate the

relationship between actual performance and post feedback reactions. Three post feedback reactions will be examined: self-efficacy, satisfaction, and fairness perception. For the individuals that accept the feedback, the relationship between feedback and reactions is anticipated to be similar to the relationships previously hypothesized between perceived performance and reactions. No relationship is expected for individuals that do not accept the feedback.

H10a: The relationship between actual performance and post feedback self-efficacy will be moderated by feedback acceptance. For individuals who accept the feedback, there will be a positive relationship between actual performance and self-efficacy. When feedback is not accepted, no relationship is expected between actual performance and self-efficacy.

H10b: The relationship between actual performance and post feedback satisfaction will be moderated by feedback acceptance. For individuals who accept the feedback, there will be a positive relationship between actual performance and satisfaction. When feedback is not accepted, no relationship is expected between actual performance and satisfaction.

H10c: The relationship between actual performance and post feedback fairness perceptions will be moderated by perceived performance and feedback acceptance. For individuals who accept the feedback, there will be a positive relationship between actual performance and fairness perceptions when perceived performance is high and a negative

relationship when perceived performance is low. When feedback is not accepted, no relationship is expected between actual performance and fairness perceptions.

In summary, because reactions to selection tests are related to a number of important organizational outcomes, reactions to adaptive testing, in addition to the psychometric properties of the test, are essential to consider when evaluating the utility of an adaptive test. The present study intends to evaluate the impact of test characteristics on examinees' reactions to adaptive testing. Specifically, as illustrated in Figure 1, characteristics of an adaptive test are hypothesized to determine the number of questions answered correctly, independent of actual performance. The number of questions answered correctly affects psychological reactions to the test through the mediating effects of perceptions of performance. In addition, the present study also examines the impact of feedback on posttest reactions. Because perceptions of performance are determined by the characteristics of the test, and not actual performance, the feedback one receives may be inconsistent with expectations. Thus, the potential exists for feedback that reflects actual performance to have an additional impact on test takers' reactions in an adaptive testing context.

Method

Development of the adaptive test

Test items. The items used to develop the test bank for the adaptive test in this study originated from three tests administered during a single semester of an Introductory Psychology class. 1167 students enrolled in two sections of Introductory Psychology

responded to 197 questions over the course of a semester. Each question was multiple-choice and consisted of four response options.

Estimation of item parameters. Prior to item calibration, the response distributions were examined. No variance was observed for six items so they were removed from the data set. The remaining 191 items were calibrated using BILOG 3.1 (Mislevy & Bock, 1989). The one, two, and three-parameter models were compared to see which model yielded superior fit. The likelihood-ratio chi-square statistic for the test of fit indicated that the three-parameter model provided superior fit for the data. Next, using the three-parameter model, the chi-square goodness-of-fit tests for individual items were examined for evidence of lack of fit. Four items displayed a significant chi-square value, $p < .01$. These four items were removed, and the remaining 187 items were re-calibrated. The chi-square statistics for these 187 items were again inspected, and no evidence for lack of fit was found. The standardized posterior residuals for individual items as well as the fit of the item characteristic curves (ICC) also were examined for the remaining 187 items for evidence of fit. Based on these procedures, the three-parameter model appeared to have the best fit and thus, was used for item calibration.

The mathematical form of the three-parameter model is

$$P_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta - b_i)}}$$

This equation expresses the probability of a correct response in terms of three parameters. The a , or discrimination, parameter is proportional to the slope of the ICC at the point of inflection. The b , or difficulty, parameter denotes the point on the ability scale where the probability of a correct response is fifty-percent. The c , or guessing parameter, denotes the lower asymptote of the ICC and represents the probability of a correct response by an examinee with infinitely low ability. Finally, D is a constant that is usually set to 1.7 so that the logistic function will approach the normal ogive function, and θ is a measure of an examinee's ability (Lord, 1980; Hambleton et al., 1991).

Ability estimation. Ability (θ) will be estimated using a conditional maximum likelihood estimation procedure. To estimate an examinee's ability, the first derivative of the log likelihood function is set equal to zero and solved for θ . Because the likelihood equation is nonlinear, it cannot be solved explicitly. For that reason, the Newton-Raphson procedure will be used. The Newton-Raphson procedure is an iterative procedure that takes the form

$$\theta_{m+1} = \theta_m - \frac{D \sum_{i=1}^n a_i (u_i - P_i) (P_i - c_i) / P_i (1 - c_i)}{D^2 \sum_{i=1}^n a_i^2 (P_i - c_i) (u_i c_i - P_i^2) Q_i / P_i^2 (1 - c_i)^2}$$

where θ_m is the m^{th} approximation to the maximum likelihood estimator θ , D is a constant equal to 1.7, a_i is the value for the discrimination parameter for the i^{th} item, c_i is the value for the guessing parameter for the i^{th} item, u_i represents the response scalar on

the i^{th} item, P_i denotes the probability of a correct response on the i^{th} item, and Q_i represents the probability of an incorrect response on the i^{th} item. If θ_m is the approximate solution for the m^{th} iteration, then θ_{m+1} represents an improved solution. The procedure continues to iterate until the difference between the m^{th} and the $(m+1)^{\text{th}}$ approximation falls below .001.

The estimation procedure described above fails to have a finite value when all responses to the administered items are correct or incorrect. Hence, after the initial item is administered, a finite maximum likelihood estimate of θ is not available. As a result, the new estimate of θ is obtained by adding 0.5 to θ for a correct response and subtracting 0.5 from θ for an incorrect response to the examinees starting ability estimate, which is determined by the experimental condition. This procedure is repeated until there is at least one correct and one incorrect response. Once this condition is satisfied, the maximum likelihood estimation procedure will be employed for estimating θ .

Samejima (1973) first noted that under certain conditions, the likelihood function might possess several local maxima. Under such conditions, the Newton-Raphson method may identify a false maximum value if the starting value is close to a local maximum. Although no explicit procedures are used to avoid the potential problem of several maxima, it should be recognized that such an occurrence requires an abnormal response pattern and a small number of items (Lord, 1980).

Aberrant responses pattern may have an additional adverse consequence as well. Hambleton and Swaminathan (1985) and Yen, Burket, and Sykes (1991) have documented that aberrant response patterns can produce infinite ability estimates when maximum likelihood estimation is used in conjunction with a three-parameter model. To eliminate this potential problem, if the estimated value of θ becomes larger than 5.0 or smaller than -5.0 during the estimation process, then the two-parameter model will be used to estimate the individual's ability, because the two-parameter model is not affected by aberrant responses (Hambleton & Swaminathan, 1985).

Stopping rule. Three stopping rules are typically available to the designer of an adaptive test. A test can be terminated when a pre-specified number of items have been administered, when a target level of measurement precision has been achieved, or when a predetermined amount of time has elapsed (Thissen & Mislevy, 1990). In the present study, the test is terminated after a pre-specified number of questions have been answered by the examinees. Every examinee will be presented exactly 35 questions regardless of the experimental condition. Pilot testing has found that by administering 35 items, the standard error of estimation [$SE(\theta)$] for most examinees in all conditions was below 0.39, which corresponds to a conventional level of reliability of 0.85.

Subjects

Participants were 162 (63 male and 99 female) undergraduate students enrolled in psychology classes. All participants volunteered for the experiment and received course credit for their participation.

Design

Participants were randomly assigned to experimental conditions. The design was a 3 (Difficulty of Initial Item: easy, medium, hard) X 3 (Difficulty of Subsequent Items: easy, medium, hard) between subjects factorial design.

Measures

Actual performance. Actual performance on the test was measured by estimating ability (θ) using the maximum likelihood estimation procedure described above.

Previous experience with adaptive tests. Two items were used to assess whether participants have had previous experience with computerized or adaptive tests. These questions asked participants if they had ever taken a computerized test or an adaptive test in the past.

The remaining measures used in this study were Likert-type questionnaires. Responses to these items ranged from *Strongly Disagree* (1) to *Strongly Agree* (7). The final structure of these scales was determined by a principal components factor analysis using a varimax rotation. An examination of the scree plot revealed an 11 factor solution. The loadings for individual items also were examined. Items loading greater than .50 on a single factor were retained. Seven items were removed because they failed to load above .50 on any factor or they loaded above .35 on multiple factors. These 11 factors explained 43.1 percent of the variance. Appendix A contains the final factor structure and loadings for the retained items.

The items were also analyzed using an oblique rotation and the factor correlation matrix was examined. However, because no correlations were in excess of .32, the original orthogonal rotation was retained (Tabachnick & Fidell, 1996). Confirmatory factor analysis was deemed inappropriate for scale development because of the small ratio of subjects to parameters being estimated. Bentler (1993) recommends that the ratio of sample size to the number of parameters to be estimated in a confirmatory factor analysis model is at the very least 5:1, but 10:1 is preferable. Given the large number of parameters to be estimated, there were an insufficient number of subjects to conduct confirmatory factor analysis.

Attributions of performance. Weiner (1985) describes four major types of attributions: luck, effort, task difficulty, and ability. Participants to indicated the extent to which each of these dimensions is responsible for their performance on the adaptive test. A separate five-item scale was used to assess each type of attribution. However, based on the results of the factor analysis, one item was removed from the attributions to ability scale.

Feedback acceptance. Feedback acceptance was assessed using a four-item scale. A typical item from this scale is "The feedback I received is an accurate evaluation of my performance."

Perceived fairness. A scale consisting of eight items, adapted from Quiñones (1995), was used to measure the perceived fairness of the adaptive test. A sample item from this scale is "The test used is an unfair test of a person's true capabilities."

Perceived performance. A five-item scale, adapted from Quiñones (1995), measured perceived performance on the test. An example item from this scale is “I performed very well on this test.”

Self-efficacy. Self-efficacy was measured using a twelve-item scale. Participants responded on a seven-point scale ranging from *Strongly Disagree* (1) to *Strongly Agree* (7) indicating expectations regarding his or her future level of performance on the adaptive test. In addition to examining overall self-efficacy, this scale was also broken into two subscales, one measuring test taking self-efficacy and one measuring psychology self-efficacy.

Satisfaction with the test. Satisfaction with the test was assessed using a three-item scale. A sample item from this scale is “I like taking this type of test.”

Test anxiety. A scale consisting of eleven items measured test anxiety. Anxiety researchers have recognized that anxiety can display itself in a variety of ways (Leary & Kowalski, 1993). Consistent with notion, three types of anxiety were measured in the present study: affective, behavioral, and social anxiety. The affective anxiety scale consisted of four items. An example item from this scale is “During the test, I got nervous”. The behavior anxiety scale measured whether subjects displayed certain physical reactions consistent with being nervous. A sample item from this four item scale is “My palms were sweaty while taking the test”. Finally, three items assessed participants’ social anxiety while taking the test. A sample item from this scale is “During the test, I was worried about what others would think about my performance.”

Test-taking motivation. Test-taking motivation was assessed using a seven-item scale adapted from Arvey et al. (1990). A typical item from this scale is "I was extremely motivated to do well on this test".

Experimental Conditions

Difficulty of initial item. The difficulty of the initial item presented to examinees was manipulated. Three different starting conditions were used: easy start, medium start and difficult start. An easy item is defined as an item in which the probability of a correct response for an individual of $\theta = -2.0$ is equal to fifty-percent. An item of moderate difficulty is an item where there exists a fifty-percent probability of a correct response when $\theta = 0.0$. Finally, a difficult item is defined as an item in which the probability of a correct response is fifty-percent when $\theta = 2.0$.

Difficulty of subsequent items. Items were selected for presentation based on a constrained form of the maximum information selection procedure (Weiss, 1982). According to the maximum information procedure, the next item presented to examinees is the one that provides more information than any other item in the test bank. The mathematical form of the information function is

$$I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$$

$I_i(\theta)$ is the information provided by item i at θ , $P'_i(\theta)$ is the derivative of $P_i(\theta)$ with respect to θ , $P_i(\theta)$ denotes the probability of a correct response on the i^{th} item at θ , and Q_i

represents the probability of an incorrect response on the i^{th} item at θ . In the present study, participants were assigned to one of three conditions in which the information function was constrained such that the probability of a correct response was either 30%, 50%, or 70% for a given level of θ . Thus, items presented provided the maximum amount of information for a given level of θ and $P(\theta)$.

Procedure

Upon arrival at the experimental laboratory, participants sat in front of a computer. Approximately ten participants were tested simultaneously. Experimental instructions were presented via the computer. The instructions explained to participants that they were about to take a test designed to assess their knowledge of psychology and that the test was to consist of 35 multiple choice questions. Participants were asked to imagine that the test they were about to take was to be used to determine their grade in a psychology class or as a selection test to get into graduate school. In an effort to increase likelihood that participants would experience anxiety during the test, participants were lead to believe that, upon completion of the test, their score was going to be shared with others currently taking the test. After the instructions were presented, participants responded to a pre-experimental questionnaire assessing their self-efficacy. Upon completion, participants had fifty minutes to complete the adaptive test. When the adaptive test terminated, participants responded to a posttest questionnaire assessing perceived fairness of the test, posttest self-efficacy, perceived performance, attributions of performance, in-test anxiety, and test-taking motivation. Participants then received

feedback regarding their actual level of performance on the adaptive test. Performance feedback was presented in terms of percentile rank of performance on the test. After receiving feedback, participants responded to additional questions assessing attitudes toward the feedback and additional attitudes regarding the test as well as demographic information including the participant's previous experience with adaptive tests.

Results

Table 1 contains descriptive statistics, inter-correlations, and reliabilities for the variables examined in this study.

Number of Items Correct

The characteristics of an adaptive test were hypothesized to determine the number of questions answered correctly (H1). Specifically, the difficulty of the initial item and the difficulty of subsequent items were hypothesized to predict the number of questions answered correctly independent of test taker ability. To test this hypothesis the number of questions answered correctly was regressed on the difficulty of the initial item, difficulty of subsequent items, the interaction of the two, and test takers' actual performance. Even after controlling for actual performance as determined by the examinee's θ estimate, the difficulty of the initial items and the difficulty of subsequent items were a significant predictor of the number of questions answered correctly. The main effect of difficulty of the initial item was statistically significant, $b = -1.26$, $t(1,157) = 5.41$, $p < .001$, as was the main effect for difficulty of subsequent items, $b = -5.23$, $t(1,157) = 21.60$, $p < .001$. Additionally, the interaction of difficulty of the initial item and the difficulty of

subsequent items was statistically significant, $b = 0.65$, $t(1,157) = 2.32$, $p = .022$, $R^2 = 0.70$. This result demonstrates that these two test characteristics, the difficulty of the initial item and the difficulty of subsequent items, interact to determine the number of questions answered correctly even after controlling for examinee ability.

To further illustrate how these two test characteristics determine the number of questions answered correctly, Table 2 contains the mean number of questions answered correctly by participants in each condition. An inspection of Table 2 reveals that as the difficulty of the initial item increased, participants answered fewer questions correctly. Similarly, as the difficulty of subsequent items increased, participants, on average, answered fewer questions correctly. These results are supportive of the hypothesized main effects.

Turning now to the interaction, Table 2 shows that participants in the easy initial items - hard subsequent items condition answered fewer questions correctly than expected. This pattern of results is likely due to the test bank not having a sufficient number of appropriate items for examinees in this condition.

Table 1

Descriptive Statistics and Correlation Coefficients for the Study Variables

	Mean	(SD)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 Actual Performance	0.74	0.22	--																
2 Number Correct	19.74	5.07	0.02	--															
3 Perceived Performance	3.02	1.01	0.14	0.24	0.84														
4 Anxiety	2.62	1.19	0.06	-0.08	-0.13	0.91													
5 Motivation	4.40	1.04	0.20	0.15	0.25	0.31	0.88												
6 Satisfaction (Time 1)	4.20	1.27	0.08	0.18	0.28	-0.12	0.25	0.77											
7 Satisfaction (Time 2)	4.22	1.35	0.17	0.17	0.30	-0.07	0.41	0.78	0.76										
8 Fairness (Time 1)	4.09	0.95	0.11	-0.04	0.12	-0.21	0.00	0.29	0.32	0.83									
9 Fairness (Time 2)	4.01	1.07	0.27	0.08	0.14	-0.11	0.17	0.31	0.43	0.76	0.88								
10 Self-efficacy (Time 1)	5.31	0.89	0.13	0.14	0.27	-0.17	0.18	0.11	0.17	0.02	0.06	0.89							
11 Self-efficacy (Time 2)	5.31	0.85	0.05	0.17	0.27	-0.24	0.18	0.11	0.14	0.05	0.08	0.85	0.89						
12 Self-efficacy (Time 3)	5.30	0.91	0.05	0.13	0.27	-0.22	0.24	0.18	0.19	0.08	0.08	0.84	0.92	0.91					
13 Attr. Perf to Luck	2.50	0.90	-0.20	-0.14	-0.17	0.10	-0.12	-0.14	-0.15	-0.30	-0.30	-0.14	-0.25	-0.20	0.74				
14 Attr. Perf to Effort	3.96	1.29	-0.05	0.01	0.12	-0.01	-0.09	-0.13	-0.17	0.16	0.10	0.00	-0.05	-0.08	-0.12	0.88			
15 Attr. Perf to Task	4.34	1.15	0.02	0.10	-0.05	0.10	0.07	-0.13	-0.14	-0.13	-0.11	0.02	0.04	0.03	0.10	0.02	0.77		
16 Attr. Perf to Ability	5.02	1.22	0.06	0.17	0.04	-0.04	0.14	0.08	0.08	0.28	0.19	0.02	0.02	0.03	-0.36	0.25	0.09	0.87	
17 Feedback Acceptance	3.94	1.40	-0.09	0.14	0.15	-0.06	0.06	0.24	0.24	0.30	0.33	0.03	0.06	0.12	-0.09	0.13	-0.14	0.10	0.87

Note: The main diagonal contains estimates of internal consistency reliability measured by Cronbach's alpha

$r > 0.15$ is significant at $p < .05$

$r > 0.20$ is significant at $p < .01$

N=162

Table 2

Mean Number of Questions Answered Correctly by Participants in Each Condition

Difficulty of Subsequent Items	Difficulty of the Initial Item			Total
	Easy	Medium	Hard	
Easy	27.1	25.0	23.6	25.2
Medium	21.3	19.8	17.5	19.5
Hard	14.5	15.9	13.5	14.6
Total	21.0	20.2	18.2	

Perceived Performance

The number of items answered correctly was hypothesized to predict perceived performance independent of actual performance (H2). To test this hypothesis perceived performance was regressed on the number of questions answered correctly, and actual performance was also entered as a covariate in the regression equation. The number of questions answered correctly significantly predicted perceptions of performance even after controlling for actual performance, $b = 0.05$, $t(1,159) = 2.98$, $p = .003$, $R^2 = 0.06$. Individuals who answered more questions correctly felt they performed better than individuals who answered fewer questions correctly after controlling for their actual level of performance. Thus, Hypothesis 2 was supported.

Motivation

Perceived performance was hypothesized to mediate the relationship between test characteristics and motivation (H3). Barron and Kenny (1986) present a procedure in which one conducts a series of regression equations to test for mediation. In the first step, perceived performance was regressed on number of items answered correctly and a significant relationship was obtained, $b = 0.05$, $t(1,160) = 3.15$, $p = .002$, $R^2 = 0.06$. In the second step, motivation was regressed on the number of items answered correctly. This relationship also was significant, $b = 0.03$, $t(1,160) = 2.10$, $p = .038$, $R^2 = 0.03$. In the final step, motivation was regressed onto both number of items answered correctly and perceived performance simultaneously. A significant effect was obtained for perceived performance, $b = 0.18$, $t(1,159) = 2.15$, $p = .033$, but the effect of number of items

answered correctly on motivation was no longer significant, $b = 0.03$, $t(1,159) = 1.54$, $p = .126$, $R^2 = 0.05$. Based on this pattern of result, Hypothesis 3 was supported. Perceived performance fully mediated the relationship between number of items answered correctly and motivation.

Test Anxiety

Perceived performance also was predicted to mediate the relationship between test characteristics and test anxiety (H4). Again, a series of regression equations were performed. The first regression equation was identical to that used in the initial step of Hypothesis 3. The second regression equation predicted participant test anxiety from the number of items answered correctly, and this regressor was not statistically significant, $b = -0.02$, $t(1,160) = 0.98$, $p = .329$, $R^2 = 0.01$. Because a majority of the participants expressed experiencing very little anxiety, the hypothesis was re-examined using only those participants who reported at least a minimal level of anxiety (above 2.5 on a 7 point scale). However, no significant results were obtained for this subset of participants. In addition to examining overall anxiety, each of the three anxiety subscales was also used as the dependent variable. Similar to the results obtained for the overall anxiety measure, no significant results were obtained for the affective, behavioral, or social anxiety subscales, $p > .05$. These statistically non significant results indicate that perceived performance does not mediate the relationship between the number of items answered correctly and test anxiety.

Satisfaction

Hypothesis 5 predicted that perceived performance would mediate the relationship between test characteristics and satisfaction with the test. The number of items answered correctly was a significant predictor of participants' satisfaction with the test, $b = 0.04$, $t(1,160) = 2.26$, $p = .025$, $R^2 = 0.03$. When satisfaction was regressed on the number of items correct and perceived performance, perceived performance was a significant predictor, $b = 0.31$, $t(1,159) = 3.19$, $p = .002$, but the effect of number correct on satisfaction became non significant, $b = 0.03$, $t(1,159) = 1.48$, $p = .141$, $R^2 = 0.09$. These results indicate that perceptions of performance fully mediated the relationship between the number of items answered correctly and satisfaction with the test.

Attributions

Perceived performance was hypothesized to mediate the relationship between test characteristics and test takers' subsequent attributions of performance, and self-efficacy was predicted to moderate this relationship (H6). To test for this and all subsequent interaction effects, variables were mean centered before entering them into the regression equations (Aiken & West, 1991). As previously shown, a statistically significant relationship exists between the number of questions answered correctly and perceived performance. Next, the moderating effects of self-efficacy on the relationship between perceived performance and four types of attributions, attributions to luck, ability, effort, and the task, were evaluated. The main effects of pre-test self-efficacy and perceived performance and the interaction of pre-test self-efficacy by perceived performance on attributions of performance to luck, ability, and effort were non significant, $p > .05$ (see

Table 3). A different pattern of results emerged for attributions of performance to the task. The main effect of perceived performance and the main effect of pre- test self-efficacy were not statistically significant, $b = -0.04$, $t(1,158) = 0.42$, $p = .657$, and $b = -0.02$, $t(1,158) = 0.15$, $p = .882$, respectively. However, the interaction of perceived performance and pre-test self-efficacy was statistically significant, $b = -0.26$, $t(1,158) = 2.63$, $p = .009$, $R^2 = 0.04$. Figure 2 illustrates the nature of this interaction. This figure, and all subsequent ones, was obtained using Aiken and West's (1991) procedure for graphing interactions using multiple moderated regression. Mean centered variables were regressed on the dependent variable and the regression equation obtained was used to generate the subsequent plot. Participants were classified as high or low if they scored one standard deviation above or below the mean on the respective measure. Participants who were high in self-efficacy were more likely to attribute performance on the test to the task when they perceived they performed poorly and less likely to attribute performance to the task when they perceived they performed well than participants low in self-efficacy. This pattern of results is consistent with Hypothesis 6. In summary, no significant relationships were found for attributions of performance to luck, effort, or ability, but pre-test self-efficacy was a significant moderator of the relationship between perceived performance and attributions of performance to the task.

Table 3

Tests of the Interaction of Self-efficacy and Perceived Performance on Attributions of Performance

	Attributions to Ability	Attributions to Effort	Attributions to Luck	Attributions to the Task
Self-efficacy	$\underline{b} = 0.06, p = .55$	$\underline{b} = 0.17, p = .11$	$\underline{b} = -0.13, p = .08$	$\underline{b} = -0.04, p = .67$
Perceived performance	$\underline{b} = 0.02, p = .83$	$\underline{b} = -.05, p = .69$	$\underline{b} = -0.11, p = .17$	$\underline{b} = -0.02, p = .88$
Self-efficacy X perceived performance	$\underline{b} = 0.11, p = .25$	$\underline{b} = 0.00, p = .99$	$\underline{b} = -0.05, p = .50$	$\underline{b} = -0.26, p = .01$

In addition to examining the moderating effects of overall self-efficacy, the effects of test taking self-efficacy and psychology self-efficacy were also examined. The results for test taking self-efficacy and psychology self-efficacy were identical to the results found for overall self-efficacy. Neither test taking self efficacy nor psychology self-efficacy were significant moderators of the relationship between perceived performance and attributions to luck, ability, or effort, $p > .05$. For attributions to the task, both test taking self-efficacy and psychology self-efficacy were significant moderators of the relationship between perceived performance and attributions, $b = -0.19$, $t(1,158) = 2.15$, $p = .033$, $R^2 = 0.04$, and $b = -0.18$, $t(1,158) = 2.22$, $p = .028$, $R^2 = 0.03$, respectively.

Self-efficacy

Hypothesis 7 predicted that attributions would moderate the relationship between perceived performance and posttest self-efficacy. As a test of this hypothesis, posttest self-efficacy was regressed on perceived performance, attributions to the task, and the interaction between the two. To control one's initial level of self-efficacy, pretest self-efficacy was also included in the regression equation. The regression weight for the main effect of perceived performance was not statistically significant, $b = -0.03$, $t(1,157) = 0.84$, $p = .402$, nor was the regression weight for the main effect of attributions of performance to the task, $b = 0.00$, $t(1,157) = 0.09$, $p = .928$. The interaction of perceived performance and attribution to the task also was non significant, $b = 0.00$, $t(1,157) = 0.08$, $p = .933$. Similarly, no statistically significant results were found when psychology self-

Figure 2

Interactions of Perceived Performance X Self-efficacy on Attributions of Performance to the Task

efficacy and test taking self-efficacy were used as the dependent variable. No support for Hypothesis 7 was found.

Fairness

Perceived performance was hypothesized to mediate the relationship between test characteristics and fairness, and self-efficacy was predicted to moderate this relationship (H8). To test for moderated mediation (Baron & Kenny, 1986) fairness was regressed on perceived performance, self-efficacy, and the interaction. The regression weight for the main effect of perceived performance was not statistically significant, $b = 0.12$, $t(1,158) = 1.52$, $p = .130$, nor was the regression weight for the main effect of self-efficacy, $b = -0.02$, $t(1,158) = 0.19$, $p = .884$. The interaction of perceived performance and self-efficacy to the task was also non significant, $b = 0.01$, $t(1,158) = 0.02$, $p = .889$, $R^2 = 0.01$. Similarly, no statistically significant results were found when psychology self-efficacy and test taking self-efficacy were used as moderators. No support for Hypothesis 8 was found.

Feedback Acceptance

The relationship between actual performance and feedback acceptance was hypothesized to be moderated by perceived performance (H9). Consistent with the hypothesis, the interaction of perceived and actual performance on feedback acceptance was statistically significant, $b = 1.86$, $t(1, 158) = 3.96$, $p < .001$, $R^2 = 0.12$. The main effect of actual performance, $b = -1.00$, $t(1,158) = 2.11$, $p < .001$, and the main effect for perceived performance, $b = 0.30$, $t(1,158) = 2.82$, $p = .002$, were both statistically

significant as well. Figure 3 illustrates the nature of this interaction. Participants who thought they performed poorly were more accepting of negative feedback than positive feedback. In contrast, participants who thought they performed well were more accepting of positive feedback than negative feedback. In other words, participants were more accepting of feedback that was consistent with perceptions of their own performance.

Post Feedback Reactions

Because perceived performance was unrelated to actual performance on the adaptive test, it was hypothesized that actual performance might impact examinees' reactions to the test after feedback was presented. However, in order for actual performance to affect reactions, participants first must have accepted the validity of the feedback. Hence, feedback acceptance was hypothesized to moderate the relationship between actual performance and post feedback reactions.

First, the relationship between actual performance and post feedback self-efficacy was examined. Post feedback self-efficacy was regressed on actual performance, feedback acceptance, and the interaction. To control for one's previous level of self-efficacy, posttest self-efficacy was included as a covariate. The main effect of actual performance was statistically non significant, $b = -0.02$, $t(1,157) = 0.05$, $p = .958$, as was the main effect of feedback acceptance, $b = 0.04$, $t(1,157) = 0.56$, $p = .574$. The interaction between feedback acceptance and actual performance was also not statistically significant, $b = 0.01$, $t(1,157) = 0.10$, $p = .922$, $R^2 = .03$. In addition to examining overall self-efficacy, psychology self-efficacy and test taking self-efficacy were also used as

Figure 3

Interaction of Perceived Performance X Actual Performance on Feedback Acceptance

dependent variables. No statistically significant results were obtained for these analyses. There was no evidence to suggest that feedback acceptance moderated the relationship between actual performance and post feedback self-efficacy.

The moderating effect of feedback acceptance on the relationship between actual performance and post feedback satisfaction was also examined. Post feedback satisfaction was regressed on actual performance, feedback acceptance, and the interaction. To control for one's previous level of satisfaction, posttest satisfaction was included as a covariate. The main effect of actual performance was statistically significant, $b = 0.96$, $t(1,157) = 3.15$, $p = .002$. In contrast, a statistically non significant main effect of feedback acceptance was observed, $b = 0.06$, $t(1,157) = 1.17$, $p = .224$. The interaction between feedback acceptance and actual performance was also statistically significant, $b = 0.50$, $t(1,157) = 2.64$, $p = .009$, $R^2 = .08$. As illustrated by Figure 4, for individuals who accept the feedback, a positive relationship exists between actual performance and post feedback satisfactions. For individuals who do not accept the feedback, no relationship exists between actual performance and post feedback satisfaction.

The effect of actual performance on post feedback fairness perceptions was evaluated. Fairness perceptions were hypothesized to be high when performance was consistent with expectations and low when inconsistencies occurred. So, perceived performance was expected to moderate the relationship between actual performance and post feedback fairness perceptions. However, this interaction was predicted to vary based

on participants' acceptance of the feedback. To test this hypothesis, regression analysis was used to evaluate the three-way interaction of actual performance X perceived performance X feedback acceptance on post feedback fairness perception was examined. Similar to previous analyses, to control for one's previous level of perceived fairness, posttest fairness was included as a covariate. The results of this analysis are contained in table 4. An inspection of Table 4 reveals that the three-way interaction effect was statistically significant, $p < .05$, $R^2 = .16$. Figure 5 illustrates the nature of significant three-way interaction. For participants who accepted the feedback, there was a positive relationship between actual performance and post feedback fairness perceptions for individuals who thought they performed well and a negative relationship for individuals who thought they performed poorly. In other words, when perceptions of performance were consistent with actual performance, fairness perceptions were enhanced. Turning our attention to participants who did not accept the feedback, there is still a positive relationship between actual performance and post feedback fairness perceptions for individuals who thought they performed well, but no relationship for individuals who thought they performed poorly.

Figure 4

Interaction of Actual Performance X Feedback Acceptance on Post Feedback Satisfaction

Table 4

Tests of the Interaction of Actual Performance X Perceived Performance X FeedbackAcceptance on Post Feedback Fairness Perceptions

Effect	<u>b</u>	<u>t</u> (1,153)	<u>p</u>
Posttest fairness	0.80	13.77	< 0.001
Actual performance	0.83	3.19	0.017
Perceived performance	0.01	0.17	0.086
Feedback acceptance	0.10	2.34	0.021
Actual performance X perceived performance	-0.15	0.49	0.623
Actual performance X feedback acceptance	0.15	0.90	0.372
Perceived performance X feedback acceptance	-0.03	0.73	0.465
Actual performance X perceived performance X feedback acceptance	-0.37	2.57	0.011

Figure 5

Interaction of Actual Performance X Perceived Performance X Feedback Acceptance on
Post Feedback Fairness Perceptions

Figure 5 (cont)

Interaction of Actual Performance X Perceived Performance X Feedback Acceptance on
Post Feedback Fairness Perceptions

Experience with Adaptive Tests

Post hoc analysis was performed to explore the effect of prior adaptive test taking experience on subsequent reactions to the test. Of particular interest was whether participants who possessed prior experience with adaptive tests were less likely to equate performance with the number of items answered correctly. A regression analysis was performed to test the moderating effects of experience with adaptive tests on the relationship between number correct and perceived performance. The results show that adaptive test taking experience was not a statistically significant moderator, $b = 0.38$, $t(1,158) = 0.81$, $p = .418$. Prior experience with adaptive tests did not affect the relationship between number correct and perceived performance.

Discussion

The results demonstrate that the design of an adaptive test can have a profound impact on the psychological experiences of test takers. The difficulty of the initial item and the difficulty of subsequent items interacted to determine the number of questions answered correctly independent of test taker ability. The number of questions answered correctly directly affected examinees' perceptions of performance, which in turn had an impact on test taker reactions to the adaptive test. Specifically, perceived performances fully mediated the relationship between number of items answered correctly and test takers' motivation and satisfaction. Individuals, who answered more questions correctly, regardless of actual performance, perceived themselves to have performed better and reported higher levels of motivation and satisfaction. The opposite was true for

individuals who answered fewer questions correctly. In addition, perceived performance also affected attributions of performance, although this relationship was moderated by pre-test self-efficacy.

Aside from evaluating posttest reactions, this study also explored the effect of feedback on test takers' reactions. Because perceived performance is unrelated to actual performance on an adaptive test, this study provided a unique opportunity to evaluate the effect of feedback that reflected actual performance but was independent of an individual's perceived performance. Consistent with past research, when the feedback matched participants' perceptions of performance, feedback acceptance was enhanced. The consistency between actual performance and perceived performance was also an important determinant of satisfaction. For people who accepted the feedback, satisfaction with the test was positively related to performance. In other words, they liked the test more when they performed well on it. Finally, the results also demonstrate that feedback acceptance moderates the relationship between actual performance and fairness perceptions. For people who accepted the feedback, the test was seen as more fair when the feedback agreed with their subjective performance evaluations.

The results concerning post feedback fairness are particularly interesting if you examine the findings in the context of current adaptive test design. A typical adaptive test is designed so the probability of a correct response is 50%. As a result, most high ability examinees will get fewer questions correct than they are accustomed, so they will have lower levels of perceived performance. However, the feedback they receive will be

positive reflecting their high level of ability. If this feedback is accepted, which is likely in a real-world testing context, then the results from the present study suggest that the discontinuity between perceived and actual performance will lead high performers to view the test as unfair (see Figure 5). So, the group organizations are most interested in selecting, high ability applicants, are the same group that perceives the typical adaptive test as most unfair.

This study only dealt with the perceived fairness of the score that participants received. However, numerous other dimensions of fairness may also be relevant when test takers ultimately evaluate the overall fairness of the selection process. In particular, perceptions of procedural fairness may be particularly important for evaluating selection tests. For example, applicants, regardless of their performance on the test, may view a test as unfair if they feel that a test should not be used for selection purposes.

The importance of maximizing test takers reactions is well documented in the literature. Empirical evidence suggests that enhancing both test taking motivation may be beneficial. Chan, Schmitt, DeShon, Clause, and Delbridge (1997) examined the relationship between performance on a cognitive ability test, reactions to the test, and subsequent performance on a parallel test. They found that test performance was positively related to test-taking motivation, and test-taking motivation affected subsequent performance on a parallel test even after controlling for initial performance. The study by Chan et. al., suggests that features of adaptive tests that impact motivation may also affect test performance on subsequent administrations.

Although the role of satisfaction was never examined by Chan et. al., the potential also exists for satisfaction to impact performance on subsequent administrations also. Future research is needed to explore this possibility. Nevertheless, other evidence exists indicating that maximizing satisfaction is also important. Smither et. al (1993) found that civil service applicants who liked a selection procedure reported higher levels of organizational attractiveness and were more willing to recommend the job to others. Linden and Parsons (1986) found that individuals who liked a selection interview were more likely to accept a job offer and had more positive attitudes about the company. Although these studies dealt with other selection techniques, a similar relationship may be expected for adaptive tests. As an initial exploration into this issue, post hoc analysis examined the relationship between satisfaction with the adaptive test and a single item that assessed participants' desire to take a similar test in the future. Satisfaction was found to be a significant predictor of test takers' desire to take an adaptive test in the future, $r = .25$, $p = .002$. Further research is necessary to evaluate the impact of satisfaction with an adaptive test on other outcomes such as organizational attractiveness and intentions to accept a job offer.

Perceptions of fairness of a selection procedure are related to a number of important outcomes as well. Smither et al. (1993) found that the perceived fairness of selection procedures was related to intentions to recommend the job to others for civil service applicants. Gilliland (1994), using undergraduates in an actual selection situation, found that the perceived fairness of a selection procedure had a positive impact on work

self-efficacy. Finally, Konovsky and Cropanzano (1991) found that the perceived fairness of an actual drug-testing program predicted job satisfaction, commitment to the organization, trust in management, turnover intentions, and job performance. In the present study, the perceived fairness of the adaptive test was significantly related to participants' desire to take a similar test in the future, $r = .29$, $p < .001$.

Aside from the benefits of adjusting adaptive tests to elicit more positive reactions, the adverse impact of these modifications on adaptive test precision and efficiency appears to be minimal. For example, Bergstrom, Lunz, and Gershon (1992) examined the effect of altering the probability of a correct response (difficulty of subsequent items). They found that altering the probability of a correct response had no impact on the estimation of examinee ability. Moreover, they noted that increasing the probability of a correct response from fifty-percent to sixty had little practical impact on test length. They found some increases in test length for tests when targeting tests to seventy percent, but the increases were most noticeable only for very long tests (over 100 items). Hence, altering the difficulty of subsequent items will do little to hinder the precision and efficiency afforded by adaptive tests. In another study, Lunz, Bergstrom, and Gershon (1994) examined the effect of modifying the difficulty of the starting item. The precision of the adaptive tests was unaffected by this modification while the efficiency of the tests was only minimally affected. Taken together, these two studies suggest that a test developer can attempt to improve test takers' reactions without sacrificing the efficiency and precision afforded by a traditional adaptive test.

Some researchers do not believe that developers of adaptive tests should attempt to maximize the psychological reactions of test takers. For instance, Wainer (1993) suggests that the efficiency of an adaptive test is of prime concern, regardless of the test's impact on reactions. Agreeably, adaptive test efficiency is important; however, as discussed above, test takers' reactions are not a trivial issue. In addition, the effect of altering an adaptive test to maximize reactions has virtually no impact on the test's efficiency. In light of both these facts, adaptive test designers should consider modifying currently used item selection algorithms in an effort to enhance reactions to the test.

Perceived performance was not a statistically significant mediator of the relationship between test characteristics and anxiety. While no true relationship may exist between test characteristics and anxiety, another possible explanation for this null result is the lack of anxiety experienced by most of the participants. Participants in this study were undergraduates who were taking the test to gain experimental credit. Their performance on the test was unrelated to any valued outcome that one might normally encounter in a real-world testing situation, such as being selected for a job or getting a grade for a class. As a result, the testing situation was unlikely to provoke much anxiety in the participants. Attempts were made to try to enhance the variance in anxiety by instructing the participants that their performance on the test would be shared with others that were taking the test. Nevertheless, these efforts may have had little or no effect on participant test anxiety because most of the participants reported little to no anxiety while taking the test. Future research should examine the hypothesized relationship in more

anxiety provoking situations in which test performance is associated with some valued outcome.

Both perceived and actual performance failed to have an impact on participants' self-efficacy. While no true relationship may exist, the null result may be attributed to fact that participants only had a single exposure to the adaptive test. Nease, Mudgett, and Quiñones (1999) examined the effect of feedback sign on participants' self-efficacy over repeated trials. They did not find any significant differences in self-efficacy across individual trials. Rather, self-efficacy was only affected after repeated exposure to negative or positive feedback. Because subjects in this study only took a single administration of the adaptive test, they were not faced with multiple successes or failures and therefore lacked the repeated exposure that might be necessary for a change in self-efficacy. Future research should evaluate changes in self-efficacy after repeated exposure to adaptive tests.

Despite the generally consistent findings, future research is needed to understand the impact of adaptive testing on test takers. Examining the impact of adaptive testing in a high stakes testing context would be particularly beneficial. Efforts should be made to establish a link between reactions and organizational outcomes such as intentions to accept a job offer. In addition, future research should explore the impact of repeated exposure to adaptive tests. In the present study, participants had only a single experience with the adaptive test. However, the potential exists for multiple exposures to have a differential impact on reactions. For example, low ability students may become less

motivated on an adaptive test regardless of perceptions of performance if they repeatedly receive negative feedback on adaptive tests over time.

Prior experience with adaptive tests can have additional implications as well. For example, in the present study, perceptions of performance, which drive subsequent reactions to the adaptive test, are a result of the number of questions answered correctly. Continued exposure to adaptive testing may decrease the strength of association between number of items correct and perceived performance as test takers become familiar with the scoring procedures employed in adaptive testing. Although there was no evidence of this in the present study, the participants were undergraduates with limited exposure to adaptive testing. Future research should explore the relationship between number of items correct and performance perceptions with populations more familiar with adaptive testing. Alternatively, test takers who possess no prior adaptive test experience, may, through some sort of informational program, learn not to equate the number of items answered correctly with performance on an adaptive test. Thereby minimizing the negative reactions seen in this study. Additional research is needed to evaluate the impact of such a program and to determine what information is needed that may eliminate potential adverse reactions.

Adaptive tests differ from more traditional tests along a number of dimensions. The present study was concerned solely with the impact of the difficulty of the initial item and subsequent items. Future research is needed to evaluate the impact of other features of adaptive testing on test takers' reactions. One feature to be studied is the

length of adaptive tests. Adaptive tests, because they are more efficient than paper-and-pencil tests, can be considerably shorter. However, test takers may have an adverse reaction to these shorter tests, especially if they receive a negative outcome as a result of their performance on the test. Examinees may feel that they were not given a sufficient opportunity to perform well if the test was excessively short. Additional research should explore these possibilities.

A more detailed evaluation of test takers' reactions would also be useful. The present study dealt with reactions to the overall test. Future research could explore reactions at the item level. For example, examinees could be asked to describe their experience while taking the tests. Such an account could provide information regarding how specific items may be related to reactions to the test. In addition, a running commentary of test takers' experience could provide insight into other aspects of the test or alternative psychological processes that may be driving reactions.

Designers of a computer adaptive test can, by manipulating these two features, control the number of questions answered correctly by examinees. This ability to control the number of questions answered correctly allows the test developer the opportunity to influence test takers' perceptions of performance and subsequent reactions to the test. By varying the characteristics of an adaptive test, these results suggest that a test developer can influence a test taker's perceptions of performance. Since these performance perceptions seem to be an important determinant of test-taking motivation, satisfaction, and fairness perceptions, they too are under the control of the test developer.

References

- Aiken, L. S. & West, S. G. (1991). Multiple regression: Testing and Interpreting interactions. Newbury Park, CA: Sage.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. Personnel Psychology, 43, 695-716.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. Psychological Review, 84, 191-215
- Bandura (1982). Self-efficacy mechanism in human agency. American Psychologist, 37, 122-147.
- Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). Self-efficacy: The exercise of control. New York, NY: W. H. Freeman & Co.
- Baron, R. M. & Kenny, D. A. (1986). The moderator mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51, 1173-1182.
- Bentler (1993) EQS: Structural equations program manual. Los Angeles: BMDP Statistical Software.
- Berstrom, B. A. & Lunz, M. E. (1999). CAT for certification and licensure. In F. Drasgow & J. B. Olsen (Eds.), Innovations in computerized assessment. Mahwah, NJ: Lawrence Erlbaum Associates.
- Berstrom, B. A., Lunz, M. E., & Gershon, R. C. (1992). Altering the level of difficulty in computer adaptive testing. Applied Measurement in Education, 5, 137-149.
- Betz, N. E. & Weiss, D. J. (1976). Psychological effects of immediate knowledge of results and adaptive ability testing. (Research Report 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometrics Methods Program.

- Chan, D., Schmitt, N., DeShon, R., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationship between race, test performance, face validity perceptions, and test-taking motivation. Journal of Applied Psychology, 82, 300-310.
- Chan, D., Schmitt, N., Jennings, D., Clause, C. S., & Delbridge, K. (1998). Applicant perceptions of test fairness: Integrating justice and self-serving bias perspectives. Paper presented at the Annual Conference for the Society for Industrial and Organizational Psychology. Dallas, TX.
- Farmer, S. J., Beehr, T. A., & Love, K. G. (1998) Effect of selection system fairness on post-selection behavior and attitudes. Paper presented at the Annual Conference for the Society for Industrial and Organizational Psychology. Dallas, TX
- Feather, N. T. & Simon, J. G. (1973). Fear of Success and causal attribution for outcome. Journal of Personality, 41, 525-542.
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational perspective. Academy of Management Review, 18 (4). 694-734.
- Gilliland, S. W. (1994). Effects of procedural and distributive justice on reactions to a selection system. Journal of Applied Psychology, 79(5). 691-701.
- Gist, M. E. & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. Academy of Management Review, 17, 183-211.
- Hambleton, R. K. & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston: Kluwer.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage Publications.
- Hambleton, R. K., Zaal, J. N., & Pieters, J. M. (1991). Computer adaptive testing: Theory, applications, and standards. In R. K. Hambleton & J. N. Zaal (Eds.), Advances in Educational and Psychological Testing: Theory and Applications. Boston: Kluwer.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. Review of Educational Research, 58, 47-77.

- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback seeking on behavior in organizations. Journal of Applied Psychology, 64, 349-371.
- Kluger, A. N. & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. Journal of Business and Psychology, 8, 3-25.
- Konovsky, M. A., & Cropanzano, R. (1991). Perceived fairness of employee drug testing as a predictor of employee attitudes and job performance. Journal of Applied Psychology, 76(5) 598-707.
- Leary, M. R. & Kowolski, R. M. (1993). The interaction anxiousness scale: Construct and criterion-related validity. Journal of Personality Assessment, 61, 136-146.
- Legg, S. M. & Buhr, D. C. (1992). Computerized adaptive testing with different groups. Educational Measurement: Issues and Practice, 11, 23-27.
- Linden, R. C. & Parsons, C. K. (1986). A field study of job applicant interview perceptions, alternative opportunities, and demographic characteristics. Personnel Psychology, 39, 109-122.
- Lonsbury, J. W., Bobrow, W., & Jansen, J. B. (1989). Attitudes toward employment testing: Scale developments, correlates, and "Known group" validation. Professional Psychology, 20, 340-349.
- Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum.
- Lunz, M. E., Bergstrom, B. A., & Gershon, R. C. (1994). Computer Adaptive Testing. International Journal of Educational Research, 6, 623-634.
- Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. Personnel Psychology, 47, 715-738.
- Mills, C. N. (1999). Development and introduction of a computer adaptive graduate records examination general test. In F. Drasgow & J. B. Olsen (Eds.), Innovations in computerized assessment. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mills, C. N. & Stocking, M. L. (1996). Practical issues in large-scale computer adaptive testing. Applied Measurement in Education, 9, 287-304.

- Mislevy, R. J. & Bock, R. D. (1989). PC-BILOG 3: Item analysis and test scoring with binary logistic models. Mooresville, IN: Scientific Software.
- Moe, K. C. & Johnson, M. F. (1988). Participants' reactions to computerized testing. Journal of Educational Computing Research, 4, 79-86.
- Nease, A. A., Mudgett, B. O., & Quiñones, M. A. (1999). Relationship among feedback sign, self-efficacy, and acceptance of performance feedback. Journal of Applied Psychology, 84, in press.
- Overton, R. C., Harms, H. J., Taylor, L. R., & Zickar, M. J. (1997) Adapting to adaptive testing. Personnel Psychology, 50, 171-185.
- Powell, Z. E. (1994). The psychological impacts of computerized adaptive testing methods. Educational Technology, , 41-47.
- Quiñones, M. A. (1995). Pretraining context effects: Training assignment as feedback. Journal of Applied Psychology, 80, 226-238.
- Rynes, S. L. & Connerly, M. L. (1993). Applicant reactions to selection procedures. Journal of Business and Psychology, 7 (3). 261-277.
- Rynes, S. L., Heneman, H. G., & Schwab, D. P. (1980). Individual reactions to organizational recruiting: A review. Personnel Psychology, 33, 529-542.
- Samejima, F. (1973). Homogeneous case of the continuous response model. Psychometrika, 38, 203-219.
- Schmidt, F. L., Urry, V. W., & Gugel, J. F. (1978). Computer assisted tailored testing: Examinee reactions and evaluations. Educational and Psychological Measurement, 38, 265-273.
- Segall, D. O. & Moreno, K. E. (1999). Development of the computerized adaptive version of the Armed Services Vocational Aptitude Battery. In F. Drasgow & J. B. Olsen (Eds.), Innovations in computerized assessment. Mahwah, NJ: Lawrence Erlbaum Associates.
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. Anxiety Research, 4, 27-41.

- Smither, J. W., Millsap, R. E., Stoffey, R. W., Reilly, R. R., & Pearlman, K. (1996). An experimental test of the influence of selection procedures on fairness perceptions attitudes about the organization and job pursuit intentions. Journal of Business and Psychology, 10, 297-318.
- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. Personnel Psychology, 46, 49-76.
- Silver, M. S., Mitchell, T. R., & Gist, M. E. (1995). Responses to successful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. Organizational Behavior and Human Decision Processes, 62, 286-299.
- Stocking, M. L. (1996). An alternative method for scoring adaptive tests. Journal of Educational and Behavioral Statistics, 21, 365-389.
- Sutton, R. E. (1997). Equity and high stakes testing: Implications for computerized testing. Equity & Excellence in Education, 30, 5-15.
- Taylor, M. S., Fisher, C. D., & Ilgen, D. R. (1984). Individuals' reactions to performance feedback in organizations: A control theory perspective. Research in Personnel and Human Resources Management, 2, 81-124.
- Thissen, D. & Mislevy, R. J. (1990). Testing algorithms. In H. Wainer (Ed.). Computer adaptive testing: A primer. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tonidandel, S. & Quiñones, M. A. (2000). Psychological reactions to adaptive testing. International Journal of Selection and Assessment, 8, 7-15.
- Vispoel, W. P., Rocklin, T. R., & Wang, T. (1994). Individual differences and test administration procedures: A comparison of fixed-item, computer-adaptive, and self-adaptive testing. Applied Measurement in Education, 7, 53-79.
- Wainer, H. (Eds). (1990). Computerized adaptive testing: A primer. Hillsdale, NJ: Erlbaum.
- Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format. Educational Measurement: Issues and Practice, 15-20.

- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. Psychological Review, 92, 548-573.
- Weiner, B., Frieze, I. H., Kubla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1971). Perceiving the causes of success and failure. Morristown: NJ: General Learning Press.
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. Applied Psychological Measurement, 6, 473-492.
- Wise, S. L. (1997). Overview of Practical Issues in a CAT Program. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Yen, W. M., Burket, G. R., & Sykes, R. C. (1991). Non-unique solutions to the likelihood equation for the three-parameter logistic model. Psychometrika, 56, 39-54.
- Zickar, M. J., Overton, R. C., Taylor, L. R., & Harms, H. J. (1999) The development of a computerized adaptive selection system for computer programmers in a financial services company. In F. Drasgow & J. B. Olsen (Eds.), Innovations in computerized assessment. Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix 1

Factor Structure and Loadings of Items

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Factor9	Factor10	Factor11
Test Anxiety											
During the test, I felt relaxed	0.82	0.00	-0.10	-0.09	0.05	-0.12	-0.07	-0.07	0.01	-0.05	-0.01
I was anxious when taking this test	0.82	-0.02	-0.05	0.04	0.08	-0.03	-0.06	-0.13	0.07	-0.02	-0.02
The test made me feel a little tense	0.82	-0.03	-0.06	0.00	0.06	-0.09	-0.03	0.01	0.02	0.11	0.16
During the test, I felt tense and jittery	0.78	-0.07	0.06	0.07	-0.04	0.03	-0.03	0.04	-0.09	0.05	-0.35
I was not nervous at all while taking this test	0.73	-0.01	-0.10	0.04	0.05	-0.09	-0.04	0.01	0.01	-0.02	0.01
While taking the test, I could feel my heart racing	0.71	-0.03	0.00	0.03	-0.18	0.11	-0.02	-0.03	-0.15	0.02	-0.34
While taking the test, I was worried about what kind of impression my score would make	0.69	-0.16	-0.18	0.16	0.03	-0.09	0.03	-0.03	0.16	-0.01	0.26
I was worried that others would not have a favorable impression of me upon learning about my performance	0.69	-0.21	-0.14	0.04	-0.01	-0.11	0.06	0.06	0.18	0.07	0.23
During the test, I was worried what others would think of my performance	0.68	-0.10	-0.16	0.18	0.01	-0.04	0.06	-0.04	0.15	-0.03	0.30
I could feel "butterflies" in my stomach while taking the test	0.65	-0.17	-0.06	-0.02	-0.13	0.15	0.03	0.02	-0.07	0.08	-0.37

My palms were sweating while taking the test **0.52** 0.06 0.08 0.07 -0.02 0.00 0.00 0.11 -0.04 0.00 -0.29

Self-efficacy

I am not sure I can ever reach a high level of performance on tests in general **0.77** 0.05 0.15 0.11 -0.03 0.05 -0.13 0.13 -0.11 -0.17

I think I can eventually reach a high level of performance on tests in general **0.74** -0.04 0.17 0.01 -0.02 0.05 -0.06 0.10 0.00 -0.26

I am not sure I can ever reach a high level of performance in psychology **0.72** 0.10 0.09 -0.02 0.12 0.03 -0.11 -0.03 -0.03 0.22

I think I can eventually reach a high level of performance in psychology **0.71** 0.01 0.03 0.04 0.08 0.00 -0.03 -0.03 -0.02 0.23

I don't feel that I am as capable of performing as well on tests as other people

-0.20 **0.68** 0.16 0.08 0.01 -0.01 0.03 0.00 0.06 -0.13 -0.16

I feel confident in my ability to do well in psychology

0.07 **0.67** -0.08 0.13 0.05 0.10 -0.15 0.12 -0.25 0.09 0.29

On the average, other people are probably much more capable of performing as well in psychology as I am

-0.08 **0.67** -0.13 -0.05 -0.01 0.10 0.01 0.03 0.03 -0.08 0.23

I don't feel that I am as capable of performing as well in psychology as other people

-0.07 **0.62** -0.02 0.03 0.02 0.15 -0.22 0.14 -0.19 0.11 0.28

I am a high performer on tests, in comparison to other people	-0.08	0.61	0.09	0.00	-0.06	0.02	0.16	0.05	0.18	-0.09	-0.32
I feel confident in my ability to do well on tests	-0.10	0.60	0.02	0.13	-0.04	0.16	0.04	-0.08	0.04	0.05	-0.34
I am a high performer in psychology, in comparison to other people	0.08	0.59	-0.23	-0.08	-0.04	0.23	-0.12	0.12	-0.19	0.06	0.17
On the average, other people are probably much more capable of performing as well on tests as I am	-0.19	0.55	0.02	-0.09	-0.10	0.01	0.15	0.03	0.09	-0.19	-0.14
Perceived Fairness											
The test used is an unfair test of a person's true capabilities	-0.10	0.07	0.72	0.08	0.25	0.07	-0.03	0.04	-0.08	-0.09	0.18
The test used was not a reliable and valid indicator of my capabilities	-0.15	-0.03	0.71	-0.02	0.14	0.08	0.06	0.18	0.00	0.02	0.07
Under the circumstances, the test was fair	-0.03	-0.01	0.70	0.01	0.10	0.03	-0.07	0.12	-0.06	-0.12	0.10
I feel other procedures should have been used to assess people	-0.04	-0.06	0.65	-0.10	-0.08	-0.04	0.34	0.03	-0.01	-0.03	-0.12
The testing procedure obtained accurate information about each person's capabilities	-0.11	-0.11	0.61	-0.03	0.17	0.04	0.05	0.20	0.01	0.19	0.00
Performance on the test was influenced by things that should not have been considered	-0.08	-0.06	0.60	0.05	0.00	0.02	0.02	-0.02	-0.09	-0.31	0.26

The test should not have been used to assess people	-0.04	0.11	0.58	0.01	-0.04	-0.04	0.18	0.07	-0.03	-0.04	-0.09
I have strong doubts that the test really measures a person's knowledge of general psychology	-0.06	0.07	0.55	-0.01	-0.20	-0.02	0.32	-0.09	0.00	-0.25	-0.01
Test-taking Motivation											
I tried my best on this test	-0.07	0.04	0.01	0.87	-0.13	0.04	0.10	0.07	0.05	-0.01	0.02
I tried to do the very best I could on this test	-0.06	0.10	0.01	0.84	-0.18	0.02	0.11	0.01	0.00	0.00	0.01
While taking this test, I concentrated and tried to do well	-0.04	0.06	-0.02	0.83	-0.08	0.04	0.12	0.02	-0.01	-0.11	-0.05
I didn't put much effort into this test	0.13	0.05	-0.06	0.70	-0.07	0.02	-0.05	-0.02	0.09	0.04	0.15
I pushed myself to work hard on this test	0.27	0.07	0.13	0.61	0.02	0.13	-0.05	0.19	0.03	-0.02	-0.08
I wanted to do well on this test	0.35	0.16	-0.14	0.52	-0.03	0.18	-0.11	0.13	0.03	-0.05	0.25
I didn't care how I did on this test	0.27	0.08	0.07	0.50	0.20	0.12	-0.04	-0.10	-0.05	-0.17	0.19
Attributions of Performance to Effort											
I performed at the level that I did because of the amount of effort I put into this test	-0.09	0.00	0.07	-0.04	0.85	0.10	0.00	-0.05	0.04	0.02	-0.01
My level of effort had nothing to do with my performance on the test	0.06	-0.06	0.07	-0.09	0.81	0.07	0.10	0.15	0.06	-0.05	-0.03
My level of effort resulted in my level of performance on the test	0.03	0.01	0.03	-0.09	0.79	0.15	0.09	0.03	-0.07	0.01	-0.16

My ability was responsible for my performance on the test	0.02	-0.05	0.06	0.12	0.16	-0.04	0.02	0.82	0.10	-0.17	0.04
My performance on the test would have been different if it wasn't for my ability	0.01	0.05	0.12	-0.13	0.04	-0.04	0.13	0.80	0.10	-0.06	-0.07
My ability caused my performance on the test	-0.07	-0.05	0.18	0.16	0.15	-0.02	-0.06	0.80	0.04	-0.13	0.04
My ability had nothing to do with my performance on the test	-0.05	0.02	0.21	0.09	0.16	0.04	0.04	0.72	0.03	-0.30	0.07
Attributions of Performance to the Task											
Characteristics of the test caused my performance	0.02	0.04	-0.07	0.07	-0.02	-0.01	-0.08	0.11	0.84	0.19	-0.07
Characteristics of the test had nothing to do with my performance	0.00	0.10	-0.05	0.00	0.00	-0.07	-0.02	0.13	0.80	-0.02	-0.03
The type of test was responsible for my performance	0.10	-0.03	-0.03	0.14	-0.02	0.06	-0.15	-0.14	0.75	0.07	-0.06
I probably would have performed differently with another type of test	0.03	-0.06	-0.06	-0.10	0.18	0.11	0.05	0.17	0.57	0.00	0.01
Attributions of Performance to Luck											
I had nothing to do with my performance on the test	-0.03	-0.18	-0.03	-0.08	-0.13	0.10	-0.13	-0.11	-0.05	0.77	-0.14
I had no control over my level of performance on the test	0.06	-0.14	-0.09	-0.01	-0.14	0.04	0.04	-0.12	0.08	0.73	0.02

My performance on the test was due to causes outside of myself that I have no influence over	0.13	-0.08	-0.37	-0.05	-0.22	-0.07	0.09	-0.16	-0.11	0.62	-0.05
My performance on the test was due to luck	-0.02	0.06	0.00	-0.05	0.13	-0.22	-0.07	-0.17	0.17	0.55	0.12
Other causes resulted in my performance on the test, luck had nothing to do with it	-0.01	0.03	-0.12	-0.05	0.18	-0.22	0.06	-0.21	0.28	0.52	0.01
Satisfaction with the Test											
This test was appealing to me	-0.07	-0.01	0.24	0.20	-0.18	0.31	0.21	0.04	-0.06	0.00	0.58
I did not like this test at all	-0.10	0.06	0.31	0.18	-0.20	0.18	0.11	0.07	-0.16	-0.09	0.55
I liked taking this test	-0.07	0.08	0.20	0.17	-0.16	0.27	0.34	0.00	-0.06	0.05	0.55