Reactions to Adaptive Testing: Effects of Test Length and Explanation

Scott Tonidandel
Davidson College

Miguel A. Quiñones
Rice University

Address Correspondence to:
Scott Tonidandel
1300-C Fairview
Houston, TX 77006

A portion of this paper was presented at the annual meeting for the Society of Industrial/Organizational Psychologists, Toronto, Canada, April 2002
Abstract

This study examined the impact of test length and explanation on test takers’ reactions to a computer-adaptive test. Participants took a computer-adaptive test that varied in length and then responded to questionnaires assessing their reactions to the test. The data show that extremely short adaptive tests were perceived by test takers to violate the opportunity to perform justice rule, and violation of this justice rule was related to subsequent fairness perceptions. Additional analyses evaluated the relationship between fairness perceptions and posttest intentions and the impact of providing explanation information to test takers regarding how adaptive tests work. Overall, the results suggest that the increased efficiency of adaptive tests has the potential to illicit adverse reactions in test takers but providing explanation information regarding how adaptive tests work may minimize these negative effects.
Reactions to Adaptive Testing: Effects of Test Length and Explanation

Unlike traditional paper-and-pencil tests that administer items in a linear fashion, adaptive tests present a select set of items deemed most appropriate for a particular individual based on his or her past responses. This strategy of item selection provides a number of advantages over conventional tests; most notably improved efficiency and precision (see Weiss, 1982, for a review). However, the novel design of adaptive tests may also produce a qualitatively different test taking experience for examinees. The purpose of this study is to examine how the structure of adaptive tests may affect test takers’ reactions. Of particular concern is the role of test length in determining reactions to adaptive tests. In addition, the potential impact of providing an explanation to examinees to familiarize them with how adaptive testing works will be explored.

Reactions to selection procedures have been shown to be related to a number of 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, and turnover intentions (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). Based on these findings, test administrators wishing to utilize adaptive selection and assessment procedures should be concerned about adverse reactions that may be elicited by this non-traditional method of testing.

Adaptive tests differ from conventional paper-and-pencil tests along a number of dimensions that have the potential to influence reactions. For example, Tonidandel and Quiñones (2000) found that test takers have more favorable reactions to tests that present the same
questions to all examinees and allow examinees to skip questions and return to them later. In a separate study, Tonidandel, Quiñones, and Adams (2002) demonstrate that item selection algorithms that affect the difficulty of items administered to examinees influence test takers reactions to an adaptive test.

**Test Length**

An additional, yet relatively unexplored characteristic of adaptive testing that may have important implications for test takers’ reactions is test length. Adaptive tests 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 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, they can be much shorter than traditional paper-and-pencil tests.

Aside from the enhanced efficiency of adaptive tests, the length of the test can also be affected by the choice of a stopping rule to terminate the test (Thissen & Mislevy, 1990; Bergstrom & Lunz, 1999). Some adaptive tests require that the same number of items be answered by all examinees. Adaptive tests can also be stopped when a certain level of measurement precision is attained. A third alternative stopping procedure requires that a prespecified level of confidence in the pass/fail decision has been achieved before the test is terminated (Bergstrom & Lunz, 1992; Kingsbury & Weiss, 1983). For example, if a cut score is in place, the test will terminate when the 95% confidence interval around an examinee’s ability estimate no longer contains the cut score.
The combination of improved efficiency and a non fixed-length stopping rule has the potential to produce adaptive tests that are drastically shorter than comparable paper-and-pencil tests and this may have a detrimental effect on examinee reactions. Relatively few studies have investigated the impact of test length on reactions to adaptive testing. Tonidandel and Quiñones (2000) found that pretest perceptions of adaptive selection tests were unrelated to test length. Vispoel et al. (1994) asked examinees to evaluate specific attributes of adaptive testing and found that participants liked the shorter length of adaptive tests. Moe and Johnson (1986) found students had favorable impressions of adaptive testing and reported a preference for adaptive tests over paper-and-pencil tests, in part, because of the shorter length of the adaptive test. These studies provide some evidence to suggest that test takers’ prefer shorter test. However, these studies fail to identify the mechanisms that may be responsible for the reactions observed. In addition, the impact of test length on reactions other than satisfaction is unclear.

Opportunity to Perform

Research in the personnel selection literature suggests that the length of a test may adversely affect reactions to the extent short tests deny test takers an opportunity to demonstrate their abilities. Gilliland (1993), in his framework of organizational justice, identifies opportunity to perform as a condition likely to affect applicants’ reactions to selection procedures. According to this justice rule, selection procedures are seen as more fair to the extent that applicants feel they have had a sufficient opportunity to express themselves. Research on voice has identified the opportunity to express oneself as a crucial variable in determining the perceived procedural fairness of an outcome (Thibaut & Walker, 1975; Avery & Quiñones, 2002). For example, Dipboye and de Pontbriand (1981) found that employees had a positive view of performance appraisal systems when they believed they had an opportunity to express their side of an issue.
In a selection context, voice can be viewed as the opportunity to demonstrate one’s knowledge, skills, and abilities in the testing situation (Gilliland, 1993; Arvey & Sackett, 1993). Thus, to the extent that the reduced length of adaptive tests elicits a perceived lack of opportunity to perform, reactions to adaptive tests could be adversely affected. Consistent with this reasoning, Bergstrom and Lunz (1999), when designing an adaptive test, chose to use a fixed-length stopping rule because they felt that candidates who failed the exam with a relatively short test might feel that they did not have sufficient opportunity to perform on the test. Despite this supposition, no empirical evaluation has been conducted to assess the extent to which the length of an adaptive test affects test takers’ perceived opportunity to perform or their subsequent reactions.

**Explanation**

Test length alone may not be solely responsible for test takers’ reactions to the test. Research on explanation suggests that providing information to workers in an effort to justify a decision can ameliorate potentially adverse reactions (Greenberg, 1990). More recently, the impact of explanation has been investigated in a selection context (Ployhart, Ryan, & Bennet, 1999; Weichman & Ryan, 2000; LaHuis, Perreult, & Ferguson, 2001). For example, Ployhart, Ryan, and Bennet (1999) found that explanation information led to increased fairness perceptions. Specifically, they found that providing participants with procedural information enhanced their perceptions of process fairness. Similarly, research by LaHuis, Perreult, & Ferguson (2001) obtained partial support for explanation information having an indirect effect on fairness perceptions by satisfying some of the procedural justice rules. Consistent with the aforementioned studies, the following hypothesis is offered:
Reactions to Adaptive Testing

H1: The effect of test length on perceived opportunity to perform will be moderated by explanation. Specifically, when participants take a long adaptive test, they will report similar levels of opportunity to perform regardless of the presence of an explanation. When participants take a short adaptive test, examinees will report higher levels of opportunity to perform when they receive an explanation than when no explanation is provided.

Fairness

As previously stated, opportunity to perform has been recognized as a potentially important determinant of the perceived fairness of selection procedures (Gilliland, 1993). However, according to Gilliland’s justice framework (1993), outcome fairness (i.e. distributive justice) will moderate the relationship between procedural rules and fairness perceptions of a selection procedure. Of particular importance seems to be the consistency between expectations of performance and actual performance on a selection procedure. Applicants will be more concerned about the fairness of the selection process if they perform worse than expected than when they perform better than expected. In support, Leung and Li (1990) found procedural justice impacted fairness perceptions only when outcomes were negative and not when outcomes were positive.

However, in order for the favorability of the outcome to affect reactions, test takers must first accept as valid the feedback that they receive regarding their performance. If test takers do not accept the feedback presented to them, it will likely have no effect on their reactions. Only when feedback is accepted can outcome favorability influence reactions. Nease, Mudgett, and Quiñones (2000) found that feedback acceptance moderated the relationship between feedback and subsequent reactions. Similarly, Tonidandel, Quiñones, and Adams (2002) found that
feedback acceptance moderated the relationship between feedback and test takers’ reactions to an adaptive test. A similar relationship is expected here.

H2: The relationship between perceived opportunity to perform and fairness perceptions will be moderated by outcome favorability and feedback acceptance. Specifically, test takers who accept the feedback will care more about procedural fairness when they perform worse than expected. Thus, the relationship between opportunity to perform and fairness will be strongest when individuals receive an unfavorable outcome. No relationship is anticipated for test takers who do not accept the feedback.

Posttest intentions

Reactions to selection procedures are imperative to consider to the extent that these reactions are related to important organizational outcomes. The perceived fairness of a selection procedure may be related to attitudes both during and after the hiring process (Gilliland, 1993). For example, Macan et. al. (1994) in a study of applicants for manufacturing positions found that the perceived fairness of a cognitive ability test battery was significantly related to organizational attractiveness and job acceptance intentions. Although evidence exists that reactions to other selection procedures may be related to these outcomes, there is no evidence that reactions to computer adaptive testing are related to subsequent attitudes and behaviors. Thus, another objective of this study is to determine if reactions to adaptive testing can predict posttest attitudes and intentions.

H3: Fairness perceptions of an adaptive test will be positively related to test takers’ posttest attitudes and intentions such as attractiveness of graduate school and desire to take a similar test in the future.
Perceived and actual performance

In general, test takers are not good at estimating their performance on an adaptive test. In separate studies, the correlation between perceived and actual performance on an adaptive test was only .07 (Powell, 1994) and .14 (Tonidandel, Quiñones, & Adams, 2002). These values are much lower than what is typically found for paper-and-pencil tests. For example, Macan et. al. (1994) observed a correlation of .40 between examinees’ perceived performance on a paper-and-pencil cognitive ability test and their actual score.

Examinees may be less accurate at gauging their performance on an adaptive test because they base their performance estimates on the number of questions answered correct. Tonidandel, Quiñones, and Adams (2002) found that the number of questions answered correct positively predicts test takers’ perceptions of performance on an adaptive test independent of their ability. However, estimating performance based on the number of questions answered correct will be inaccurate because of the novel scoring procedure used by adaptive tests. Unlike traditional paper-and-pencil tests that determine an examinee’s score based on the number of questions answered correctly, performance on an adaptive tests is determined by which questions are answered correctly. Thus, two test takers who get the same number of questions correct on an adaptive test can receive vastly different scores if one answers difficult questions and the other answers easier questions.

The inability of test takers to accurately gauge their performance on an adaptive test is important because the inconsistency between perceived and actual performance on an adaptive test was shown to be related to test takers’ perceptions of the fairness of an adaptive test (Tonidandel, Quiñones, & Adams, 2002). As a result, a final goal of the present study is to explore the effectiveness of an intervention designed to improve the accuracy of test takers’
perceptions of performance on an adaptive test. Specifically, the impact of providing test takers with an explanation of how an adaptive test works on the accuracy of performance perceptions will be evaluated.

H4: Explanation will moderate the relationship between perceived and actual performance such that individuals who receive an explanation regarding how adaptive tests work will be more accurate at gauging their true level of performance on an adaptive test.

Method

Participants

Participants were 137 (46 male and 91 female) undergraduate students enrolled in psychology classes. All participants volunteered for the experiment and received course credit for their participation. Preliminary data screening revealed computational difficulties in estimating ability ($\theta$) for some individuals. Seventeen subjects, whose final ability estimate failed to converge using the 3-parameter model, were removed because they received atypical estimates of theta.

Design

The design was a 2 (test length: short vs. long) X 2 (explanation: no explanation vs. explanation) between subjects factorial design.

Development of the adaptive test

The adaptive test administered in this study utilized an item bank consisting of 187 multiple-choice items with four response options. These items, assessing knowledge of general psychology, originated from three tests administered during a single semester of an Introductory Psychology class and were calibrated specifying a three-parameter model using BILOG 3.1 on
data obtained from an initial sample of 1167 students. A more detailed description of the item calibration and the development of the adaptive test can be found in Tonidandel, Quiñones, and Adams (2002).

**Measures**

**Actual performance.** Actual performance on the test was measured by estimating ability $\theta$ using a maximum likelihood estimation procedure. Theta was then converted into percentile rank of performance on the test.

**Outcome favorability.** After completing the test, participants were asked to rate their expected performance on the test by estimating their expected percentile rank. Outcome favorability was determined by comparing participants’ expected percentile rank of performance with participants’ actual percentile rank.

The remaining measures used in this study were Likert-type questionnaires consisting of seven response options ranging from *Strongly Disagree* (1) to *Strongly Agree* (7). A description of each scale follows.

**Attractiveness to graduate school.** A scale consisting of 4 questions was used to measure this construct. A sample item from this scale is “I would be less likely to apply to graduate school in psychology if I have to take a test like this one in order to get in.” Coefficient alpha for this scale was .84.

**Feedback acceptance.** Feedback acceptance was assessed using a four-item scale. A sample item from this scale is “The feedback I received is an accurate evaluation of my performance.” This scale had an internal consistency reliability of .86.
Perceived fairness. A scale consisting of nine items was used to measure the perceived fairness of the adaptive test. A sample item from this scale is “The test is an unfair test of a person's true capabilities.” Coefficient alpha for this scale was .80.

Perceived opportunity to perform. Opportunity to perform was assessed using the following two items: “A longer test would have allowed me to do better” and “There were not really enough questions to allow me to get a good score.” The internal consistency reliability for this scale was .82.

Perceived performance. A scale consisting of four items was used to assess perceptions of performance on the test. A sample item from this scale was “I know I did better than most people on this test.” Coefficient alpha for this scale was .92.

Willingness to take a similar test in the future. A sample item from the four-item scale used to measure this construct is “I would not want to take a similar test in the future.” Coefficient alpha for this scale was .89.

Experimental Conditions

Test length. Test length was manipulated by varying the stopping rule used to terminate the test. In the present study, the test was terminated when a target level of measurement precision has been achieved. For the short test condition, the test was terminated when the standard error of estimation \(SE(q)\) fell below .548, which corresponds to a conventional level of reliability of .70. For the long test condition, a \(SE(q)\) of less than .316, corresponding to a reliability of .90, was necessary to stop the test. Because of time constraints, the test was further limited to a maximum of 40 items regardless of condition.

Explanation. Participants in the explanation condition received the following additional set of instructions describing the test they were about to take:
The test you are about to take is "computer adaptive." This means that as you answer each question, the computer uses information about how you answered that question and how you answered previous questions to determine which question you get next. The computer adaptive test targets questions to your individual ability level. This means that you won't answer questions that are too easy or too difficult for you. Since the questions you receive are tailored to your ability level, you may not receive the same questions as someone else. As a result, computer adaptive tests are scored differently than most paper-and-pencil tests. Instead of basing your score on the number of questions answered correctly, your score will be also be determined by the level of difficulty of those questions answered correctly.

Participants in the no explanation condition did not receive any of this additional information.

Procedure

Between 2 and 8 individuals participated in the experiment simultaneously. Participants were seated in front of a computer and all instructions were presented via the computer. Participants were told they were about to take a test assessing their knowledge of psychology. Participants were asked to imagine that the test they were about to take was going to be used to determine admission into graduate school. For individuals in the explanation condition, additional information containing a description of the adaptive test was imbedded within these initial instructions. After the instructions were presented, a preliminary questionnaire was administered. When the adaptive test terminated, participants responded to a posttest questionnaire assessing perceptions of performance and perceived opportunity to perform. After completing this questionnaire, actual performance feedback conveyed in terms of percentile rank
of performance was presented to participants. Participants then responded to additional questions assessing reactions to the test and demographic information.

Results

Manipulation check. The effectiveness of manipulating the stopping rule to affect test length was examined. On average, participants in the short test condition answered 13.6 questions (SD = 8.4), while participants in the long test condition received average of 33.3 questions (SD = 8.3). The difference between these means was statistically significant, \( F = 167.82, p < .001 \). Thus, our manipulation of the stopping rule was sufficient to affect the length of the test.

Descriptive statistics, inter-correlations, and reliabilities for the variables examined in this study are contained in Table 1. Because actual performance was correlated with the experimental conditions, performance on the test was used as a control variable in all of the analyses.

Perceived opportunity to perform. Explanation was hypothesized to moderate the relationship between test length and perceived opportunity to perform (H1). As a test of this hypothesis, a Test Length ¥ Explanation ANOVA on opportunity to perform was computed, and performance on the test was used as a covariate. The main effect for test length, \( F(1,116) = 21.88, p < .001 \), and the interaction between test length and explanation, \( F(1,116) = 6.04, p = .02 \), were statistically significant, while the main effect for explanation was not significant, \( F(1,116) = 1.43, p = .23 \). The interaction of Test Length ¥ Explanation accounted for an additional 4% of the variance in opportunity to perform (\( R^2 = .04 \)). Figure 1 illustrates the nature of this interaction. Participants in the long test condition reported similar levels of opportunity to perform regardless of whether or not an explanation was provided. For participants in the short
test condition, the presence of explanation information enhanced perceived opportunity to perform. Thus, hypothesis 1 was supported.

**Fairness.** Outcome favorability and feedback acceptance were expected to moderate the relationship between opportunity to perform and fairness perceptions (H2). To test this hypothesis, fairness perceptions were regressed on opportunity to perform, favorability of the outcome, feedback acceptance, and their interactions. Test length, explanation, and their interaction were included as covariates. The three-way interaction effect of Opportunity to Perform × Favorability of the Outcome × Feedback Acceptance was statistically significant, $B = -0.31$, $t(109) = 5.81$, $p = .02$ and accounted for an additional 4% of the variance ($R^2 = .04$).

Figure 2 illustrates the nature of the significant three-way interaction. As shown in figure 2, for participants who did not accept the feedback, there was a positive relationship between opportunity to perform and fairness perceptions when outcome favorability was positive and a negative relationship when outcome favorability was negative. For participants who accepted the feedback, there was no relationship between opportunity to perform and fairness perceptions when outcome favorability was positive and a positive relationship when outcome favorability was negative. In other words, for participants who accepted the feedback, violations of the opportunity to perform justice rule had a larger impact on fairness perceptions when individuals performed worse than expected. This pattern of results is consistent with the predictions of hypothesis 2.

**Posttest intentions.** Fairness perceptions were hypothesized to positively predict attractiveness of graduate school and willingness to take a similar test in the future. (H3). In order to control for the influence of other study variables, hierarchical regression analysis was used. Explanation, number of items, actual performance and opportunity to perform, were
entered as control variables in the initial step of the regression. Fairness perceptions were then entered separately in a second step. The results from these analyses are contained in Table 2. Fairness significantly predicted both attractiveness to graduate school, $b = 0.44$, $t(1,115) = 3.27$, $p < .001$, and willingness to take a similar test in the future, $b = 0.67$, $t(1,115) = 5.82$, $p < .01$, even after controlling for other study variables, thus providing support for Hypothesis 3.

**Perceived and actual performance.** Explanation was predicted to influence the relationship between perceived and actual performance on the task. To test this hypothesis, we examined the correlation between perceived and actual performance for individuals who did and did not receive an explanation regarding how adaptive tests work. After controlling for the effects of test length, the correlation between perceived and actual performance was $r = .19$, $p = .16$, for individuals in the no explanation condition and $r = .50$, $p < .001$, for individuals in the explanation condition. These two correlations were statistically significantly different from one another, $p = .04$, indicating that, as predicted by hypothesis 4, participants were much more accurate at estimating their actual performance when they were provided with an explanation regarding how an adaptive test works than participants in the no explanation condition.

**Discussion**

Adaptive tests, due to their increased efficiency over traditional paper-and-pencil tests, hold great promise for reducing testing time. However, the results from the present study indicate that this reduction in test time has the potential to elicit adverse reactions in test takers. Specifically, extremely short adaptive tests were perceived by test takers to violate the opportunity to perform justice rule, and violation of this justice rule was related to subsequent fairness perceptions. The relationship between opportunity to perform and fairness was moderated by both outcome favorability and feedback acceptance. When feedback was accepted,
examinees’ fairness perceptions were more strongly related to violations of the opportunity to perform justice rule when outcomes were unfavorable than when outcomes were favorable. In other words, test takers seem care more about not having an opportunity to perform when they perform worse than anticipated.

The present study also provides some evidence to suggest that interventions can be designed to ameliorate some of the negative effects of extremely short tests. Providing test takers with an explanation to familiarize them with adaptive tests tended to increase perceived opportunity to perform for participants in the short test conditions. These results are particularly interesting given that the explanation manipulation never mentioned that adaptive tests were more efficient and shorter than traditional tests. Future research is needed to document the specific aspects of explanation that are responsible for the effects observed in the present study. In addition, the psychological mechanisms that can account for the positive effects of explanation need to be identified. Future research should also explore whether explanation can ameliorate some of the negative reactions elicited by other aspects of adaptive testing aside from test length such as the difficulty of items presented to test takers and the ability to navigate within the test (Tonidandel & Quiñones, 2000; Tonidandel, Quiñones, & Adams, 2002).

From an organizational perspective, reactions to selection procedures are of concern to the extent they impact important organizational outcomes such as attractiveness to the organization and willingness to apply for the job in the future. This study provides some initial support that reactions to adaptive testing may be related to these important organizational outcomes. In the present study, fairness perceptions positively predicted both attractiveness to graduate school and desire to take a similar test in the future. Although fairness perceptions of an adaptive test were related to subsequent attitudes, future research is needed to assess whether
reactions to adaptive testing are also indicative of examinees’ posttest behavior. In order to examine important posttest behaviors, future research should examine reactions to adaptive testing and subsequent attitudes and behaviors in an actual selection context.

These results also shed some light on the relationship between perceived and actual performance on adaptive tests. Past research has identified perceptions of performance as a crucial variable in determining examinees’ reactions to an adaptive test (Tonidandel, Quiñones, & Adams, 2002). In the present study, the presence of explanation information regarding how adaptive tests work enhanced the accuracy of test takers’ perceptions of performance. Nevertheless, the mechanisms responsible for the increased accuracy of participants in the explanation condition remain unknown. More research is needed to evaluate the effects of explanation information on the accuracy of performance perceptions with the goal of identifying the specific aspects of explanation information and the underlying psychological processes responsible for these effects.

Based on the results from the present study, caution should be used when designing an adaptive test with a variable-length stopping rule. Adaptive tests designed in this manner have the potential to be much shorter than test takers are used to and thus, may be more likely to elicit more negative perceptions of fairness. Nevertheless, the presence of information to familiarize test takers with adaptive testing can minimize these negative effects.
References


Table 1

Descriptive Statistics and Correlation Coefficients for the Study Variables

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Explanation</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Number of items</td>
<td>0.04</td>
<td>0.25</td>
<td>-0.21*</td>
<td>-0.37**</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Outcome favorability</td>
<td>4.06</td>
<td>1.40</td>
<td>0.21*</td>
<td>-0.27**</td>
<td>0.08</td>
<td>.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Opportunity to perform</td>
<td>4.38</td>
<td>1.36</td>
<td>0.20*</td>
<td>0.23**</td>
<td>-0.31**</td>
<td>0.11</td>
<td>(.86)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Feedback acceptance</td>
<td>3.52</td>
<td>0.86</td>
<td>0.23*</td>
<td>-0.26**</td>
<td>0.12</td>
<td>0.38**</td>
<td>.19*</td>
<td>(.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Perceived fairness</td>
<td>4.85</td>
<td>1.25</td>
<td>0.04</td>
<td>0.01</td>
<td>0.06</td>
<td>-0.16</td>
<td>0.01</td>
<td>0.24**</td>
<td>(.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attractiveness of</td>
<td>3.80</td>
<td>1.16</td>
<td>0.14</td>
<td>-0.12</td>
<td>0.05</td>
<td>0.03</td>
<td>0.06</td>
<td>0.49**</td>
<td>0.55**</td>
<td>(.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>graduate school</td>
<td>0.65</td>
<td>0.22</td>
<td>-0.26**</td>
<td>-0.24**</td>
<td>0.61**</td>
<td>-0.35**</td>
<td>-0.43**</td>
<td>0.01</td>
<td>0.37**</td>
<td>0.29**</td>
<td>(.90)</td>
<td></td>
</tr>
<tr>
<td>Willingness to take test</td>
<td>3.78</td>
<td>1.27</td>
<td>-0.12</td>
<td>0.11</td>
<td>-0.38**</td>
<td>-0.45**</td>
<td>-0.11</td>
<td>-0.11</td>
<td>0.34**</td>
<td>0.31**</td>
<td>0.36**</td>
<td>(.92)</td>
</tr>
</tbody>
</table>

* denotes p < .05
** denotes p < .01

a coded: 0=no explanation, 1=explanation
b coded: 0=long, 1=short

N = 120
Table 2

Regression Results for Attractiveness of Graduate School and Willingness to Take Test in Future

<table>
<thead>
<tr>
<th>Variable</th>
<th>Willingness to take test in future</th>
<th></th>
<th>Attractiveness of graduate school</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B step 1</td>
<td>B step 2</td>
<td>B step 1</td>
<td>B step 2</td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explanation</td>
<td>0.49*</td>
<td>0.27</td>
<td>0.39</td>
<td>0.25</td>
</tr>
<tr>
<td>Number of items</td>
<td>0.02</td>
<td>0.16</td>
<td>0.28</td>
<td>0.37</td>
</tr>
<tr>
<td>Actual performance</td>
<td>2.00**</td>
<td>1.60**</td>
<td>2.38**</td>
<td>2.12**</td>
</tr>
<tr>
<td>Opportunity to perform</td>
<td>0.10</td>
<td>-0.05</td>
<td>-0.01</td>
<td>-0.11</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairness</td>
<td></td>
<td>0.67**</td>
<td></td>
<td>0.44**</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.11**</td>
<td>0.31**</td>
<td>0.14**</td>
<td>0.20**</td>
</tr>
<tr>
<td>△$R^2$</td>
<td></td>
<td>0.20**</td>
<td></td>
<td>0.06</td>
</tr>
</tbody>
</table>

* denotes $p < .05$

** denotes $p < .01$
Figure 1. Interaction of explanation with test length on perceived opportunity to perform.
Figure 2. Interaction of perceived opportunity to perform and outcome favorability with feedback acceptance on fairness perceptions.