DOCUMENT RESUME

ED 371 007 TM 021 638

AUTHOR Wise, Steven L.; And Others
TITLE Comparing Computerized Adaptive and Self-Adapted Tests: The Influence of Examinee Achievement Locus of Control.
PUB DATE Apr 94
NOTE 22p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 5-7, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Adaptive Testing; Anxiety; Attribution Theory; *College Students; Comparative Analysis; *Computer Assisted Testing; Higher Education; *Locus of Control; Personal Autonomy; Pretests Posttests; Self Concept
IDENTIFIERS *Self Adapted Testing

ABSTRACT

This study investigated the relationship between examinee achievement-specific locus of control and the differences between self-adapted testing (SAT) and computerized adaptive testing (CAT) in terms of mean estimated proficiency and posttest state anxiety. Subjects were 379 college students. A disordinal interaction was found between test type and locus of control. Examinees with an internal locus of control were affected positively by the SAT (relative to the CAT). For examinees with an external locus of control, however, the SAT appeared to have a negative effect on both estimated proficiency and posttest state anxiety. There are four tables and two figures. (Contains 18 references.) (Author/SLD)

Comparing Computerized Adaptive and Self-Adapted Tests: The Influence of Examinee Achievement Locus of Control

Steven L. Wise, Linda L. Roos, Barbara S. Plake

University of Nebraska-Lincoln

Paper Presented at the 1994 Annual Meeting of the National Council on Measurement in Education, New Orleans

Abstract

This study investigated the relationship between examinee achievement-specific locus of control a.id the differences between self-adapted testing (SAT) and computerized ad ptive testing (CAT) in terms of mean estimated proficiency and posttest state anxiety. A disordinal interaction was found between test type and locus of control. Examinees with an internal locus of control were affected positively by the SAT (relative to the CAT). For examinees with an external locus of control, however, the SAT appeared to have a negative effect on both estimated proficiency and posttest state anxiety.

Comparing Computerized Adaptive and Self-Adapted Tests: The Influence of Examinee Achievement Locus of Control

During the past several years, there have been a number of investigations into the effects and dynamics of self-adapted testing (SAT), a type of computerized testing that was introduced by Rocklin and O'Donnell (1987). SAT represents a novel application of item response theory (IRT) in which examinees are allowed to choose the difficulty levels of their test items. This is in contrast to the more common computerized adaptive test (CAT), in which a computer algorithm is used to choose the item difficulties for an examinee based on that examinee's responses to previously administered test items. An overview of the basic principles of SAT is found in Rocklin (in press).

According to the invariance principle of IRT, the expected proficiency estimate for an examinee should be the same regardless of whether a CAT or a SAT is administered. Rocklin and O'Donnell (1987) found, however, that examinees who were administered a SAT obtained significantly higher mean proficiency estimates than examinees who were administered conventional computer-based tests. Several subsequent studies have compared SAT with CAT, finding that examinees receiving a SAT obtained significantly higher mean proficiency estimates (Wise, Plake, Johnson, & Roos, 1992; Roos, Plake, & Wise, 1992; Vispoel & Coffman, in press). These investigations of SAT have used test item pools from a variety of content areas and have employed an array of IRT calibration and scoring models.

The difference in mean estimated proficiency between SAT and CAT has been found to interact with other variables. Several studies have found significant interactions between test type and examinee scores on the Test Anxiety Inventory (Spielberger, 1980), with the difference in mean estimated

proficiency between SAT and CAT increasing with examinee test anxiety (Rocklin & O'Donnell, 1991; Vispoel & Coffman, in press; Vispoel, Rocklin, & Wang, in press; Vispoel, Wang, de la Torre, Bleiler, & Dings, 1992). In addition, Vispoel et al. (in press) found a significant interaction between examinee verbal self concept and test type, with the largest difference in mean estimated proficiency between SAT and CAT being associated with low examinee verbal self concept.

Use of a SAT appears to moderate the influence of examinee anxiety on test performance. Two studies comparing SAT and CAT have found that examinees administered a SAT report significantly lower posttest state anxiety than examinees administered a CAT (Roos et al. 1992; Wise et al. 1992). In addition, several researchers have found that a SAT yields proficiency estimates that are less related to test anxiety than those obtained when a CAT or a conventional test is used (Rocklin & O'Donnell, 1991; Vispoel & Coffman, in press; Vispoel et al., in press; Vispoel et al., 1992). These researchers have concluded that use of a SAT reduces the influence of anxiety on proficiency estimates.

Although a number of investigations of SAT have yielded similar findings, it remains unclear why SAT has a positive effect on examinees. There have been several hypotheses proposed. The first, a self-monitoring hypothesis, was discussed by Rocklin and O'Donnell (1987). They suggested that "an examinee has access to a variety of information (including current affective and motivational states) relevant to optimal item selection beyond that which would be available to a traditional computerized testing algorithm" (p. 318). Presumably, examinees use item difficulty selection as a means to strategically adjust their levels of anxiety and motivation to levels that promote maximal test performance.

A second hypothesis was suggested by Wise (1992) and Vispoel et al. (in press). Research has indicated that test-anxious examinees experience test-irrelevant thoughts that interfere with test performance (Sarason, 1986). In SAT, the act of continually choosing item difficulty levels may block an examinee's test-irrelevant thoughts and, consequently, keep the examinee more focused on the test.

Wise (in press) proposed a third hypothesis regarding the effects observed when a SAT is administered that is based on research drawn primarily from the psychological literature. This research has focused on the construct of *perceived control*. Numerous studies have indicated that, in a stressful situation, if people believe that they have some control over the source of the stress, they exhibit improved performance on cognitive tasks, lower anxiety, and increased motivation. Assuming that the testing situation is stressful and examinees who are given an opportunity to choose item difficulty levels perceive that they have control over the stressful situation, the results found in previous investigations of SAT can be explained by this hypothesis.

This study investigated the tenability of the perceived control hypothesis. In many contexts an important individual difference construct is *locus of control*—the degree to which people feel that they generally have control over various aspects of their lives (Lefcourt, 1982). In a testing situation, examinees varying in perceptions of locus of control might respond differently to the control provided by a SAT. If perceived control is integral to the higher test performance yielded by a SAT, then it is reasonable to expect that being allowed to choose one's item difficulty levels will most positively affect examinees who believe that they have a great deal of control over their test performance (an internal orientation). Examinees who believe that they have relatively little control over their test

performance (an external orientation) would be expected to be less positively affected by a SAT.

In this study, an achievement-specific measure of locus of control was used. The following three research questions were investigated:

1. Does the difference between a SAT and a CAT in estimated proficiency vary across levels of examinee achievement-specific locus of control? It was expected that the SAT-CAT difference in estimated proficiency would be greatest for those examinees indicating an internal locus of control.
2. Does the difference between a SAT and a CAT in mean posttest anxiety vary across levels of examinee achievement-specific locus of control? It was expected that the SAT-CAT posttest anxiety difference would be greatest for those examinees reporting an internal locus of control.
3. For examinees administered a SAT, are the choices of item difficulty levels related to level of achievement-specific locus of control? It was expected that examinees reporting an external locus of control would perceive that the SAT did not provide control over the testing situation and would, therefore, exhibit less variability in their difficulty level choices than examinees reporting an internal locus of control.

Method

The experimental design used in this study was largely a replication of the Wise et al. (1992) comparison of SAT and CAT. The primary difference in this study was the measurement of examinee achievement-specific locus of control.

Examinees

A total of 379 students from a large midwestern university participated in the study. The students were enrolled in an introductory statistical methods course. Data were collected from 11 course sections during two summer sessions and the fall semester of 1993. The group of subjects consisted of 128 males and

251 females. There were 268 undergraduates and 111 graduate students in the sample. Examinees were randomly assigned to the two experimental conditions used in the study.

During the early part of the course, students are routinely tested to assess their working familiarity with the types of algebra skills necessary in introductory statistics. Students who obtain low scores on this test are required to attend a one-hour review session concerning relevant algebra skills.

Instruments

The primary instrument used was a computer-based algebra test administered via IBM microcomputers using MicroCAT testing software (Assessment Systems Corporation, 1988). Each examinee received a 20-item test drawn from a pool of 91 multiple-choice items. Both SAT and CAT versions of the algebra test were developed. Examinee proficiency was estimated using a maximum likelihood method. Wise et al. (1992) contains detailed information regarding item pool development, IRT model fit, and test instructions for the SAT and CAT versions.

Achievement-specific locus of control was measured using the Achievement subscale of the Multidimensional-Multiattributional Causality Scale (MMCS; Lefcourt, Von Baeyer, Ware, & Cox, 1979). This scale consists of 24 statements to which examinees respond using a 5-point scale. An examinee's total achievement locus of control was computed by forming the difference between the summed scores for the external items and the summed scores for internal items. The resultant achievement locus of control scores could range from -48 to +48, with negative scores indicating examinees who reported predominately internal attributions of control over their achievement (ability and effort) and positive scores indicating examinees who reported predominately external attributions (test difficulty and luck).

There were two additional instruments used in this study. The Test Anxiety Inventory (TAI) was used to measure an examinee's general feelings of anxiety regarding tests. In addition, the State Anxiety Scale of the State-Trait Anxiety Inventory (Spielberger, Gorsuch, &: Lushene, 1970) was used as a measure of situation-specific anxiety both before and after the testing session.

Procedure

Testing was conducted during the first two days of each of the five-week summer sessions and during the first week of the fall semester of 1993. During the first class meeting the students were informed that the test scores would be used to identify students in need of an algebra review. At this time the students (a) completed demographic sheets, (b) signed up for a time to be administered the algebra test, (c) and completed the TAI and the achievement subscale of the MMCS.

The algebra test was administered in a large quiet room containing 12 IBM PS/2 Model 55 microcomputers. Half of the computers were randomly assigned to administer a SAT, with the remainder administering a CAT. At several points during testing, the computers were randomly re-assigned to the SAT and CAT conditions.

Students were tested in groups ranging in size from 1 to 12. As each student arrived for testing, the test administrator directed the student to choose a computer. By their choice of randomly designated computers, students assigned themselves to an experimental condition. After being seated at a computer, each student completed the State Anxiety Scale (in a paper & pencil form) and began the algebra test. No time limit was imposed, and pencils and scratch paper were provided. Immediately after completion of the test, each student was again administered the State Anxiety Scale and informed, via the computer, whether he or she would be required to attend an algebra review session.

Data Analysis

The first two research questions focused on comparisons between SAT and CAT in terms of estimated proficiency and posttest state anxiety. Three-factor analyses of variance were used for each dependent variable. The primary independent variable was test type (SAT, CAT). A second independent variable was locus of control (internal, external); examinees with negative achievement locus of control scores were classified as internals whereas examinees with positive scores were classified as externals. In addition, test anxiety was used as a blocking variable because it had been found to interact with test type in previous SAT studies. Two levels of test anxiety were formed using a median split of the TAI score distribution; examinees with scores less than 39 were classified as low in test anxiety, with the remainder classified as high in test anxiety.

The third research question focused on the difficulty level choices made by the examinees in the SAT condition. The number of times that each examinee changed difficulty levels during the first 15 items was computed and used as the dependent variable. The analysis used the data from the first 15 items because that was how many items were contained in most of the difficulty levels. Hence, an examinee had unconstrained choices among the six difficulty levels for only the first 15 items. Because it was expected that the distribution of difficulty changes would be markedly skewed, the large scale approximation of the Mann-Whitney *U* test was used to compare the locus of control groups. A .05 level of significance was used in all data analyses.

Results

The results of the analysis of variance for estimated proficiency are found in Table 1. There was a significant interaction between test type and locus of control. In addition, there was a significant main effect for test anxiety which

was expected due to its role as a blocking variable in the analysis. To understand the nature of the significant interaction, tests of simple main effects were performed for test type at each locus of control group. The results of these tests, which are also contained in Table 1, indicate that the test type effect was significant only for the externals. A graph of the significant interaction is shown in Figure 1. Descriptive statistics for estimated proficiency, broken down by test type and locus of control are found in Table 2. Note that the sample contained only 34 examinees (9%) who reported an external locus of control.

Table 3 contains the results of the analysis of variance for posttest state anxiety. Consistent with the results for estimated proficiency, there was a significant main effect for test anxiety and a test type by locus of control interaction. In addition, there was a significant main effect for test type. A graph of the test type by locus of control interaction is shown in Figure 2. To follow up the significant interaction, tests of simple main effects were performed for test type at each locus of control group. The simple effects were found significant for each locus of control group, although the direction of the anxiety differences between the test types differed across the locus of control groups. Descriptive statistics for posttest state anxiety are found in Table 4.

Results of the analysis of difficulty level changes for the SAT group revealed that the locus of control groups exhibited no significant difference in their use of the SAT ($z = -0.59$, $p = .553$). For the internals the number of difficulty level changes ranged from 0 to 12, with a median of 4.0. For the externals, the number of difficulty level changes ranged from 0 to 11, with a median of 3.5.

Discussion

On the basis of the perceived control hypothesis, it was expected that the differences in estimated proficiency and posttest state anxiety between SAT and CAT would vary across levels of examinee achievement locus of control. More

specifically, it was expected that the difference between SAT and CAT found in many prior studies (higher test performance, lower posttest anxiety) would be larger for internals than for externals.

Significant interactions were found between test type and locus of control for both estimated proficiency and posttest state anxiety, indicating that an examinee's achievement locus of control moderates the effect of a SAT. For internals, the SAT yielded lower posttest anxiety and higher test performance (though not significantly higher than that yielded by the CAT). For externals, however, the SAT yielded *higher* posttest anxiety and *lower* test performance. This latter finding was unanticipated, and it represents the first empirical evidence that SAT may have a negative effect on some examinees.

How can the disordinal interactions between test type and achievement locus of control be explained? There may be differences between the locus of control groups in the desirability of control. Examinees with an external locus of control, who feel that they have relatively little control over their test performance, appear to react negatively to the control provided by a SAT. Having greater control implies a greater responsibility for one's test performance, and because an examinee administered a SAT chooses the difficulty levels of the items, he or she cannot readily attribute poor test performance to test difficulty. This suggests that externals may not desire the control provided by a SAT, because it eliminates a source of an external attribution for one's test performance.

The hypothesis that, in the current study, externals did not view increased perceived control as desirable is speculative and warrants further investigation. There is, however, some empirical support for the idea that persons do not always see increased perceived control as beneficial. Burger (1989) discussed conditions under which persons can have negative reactions to increased

perceived control. Moreover, Baron and Logan (1993) reviewed a series of studies indicating that the effects of increased perceived control are moderated by desire for control; positive effects were consistently found only for subjects with a high desire for control.

The analysis of examinee difficulty level choices found that internals and externals made approximately the same number of difficulty level changes. This indicates that the degree to which an examinee becomes engaged in the process of choosing difficulty levels is not moderated by achievement locus of control. Inspection of the difficulty level choices made by examinees, however, suggests that a variety of difficulty selection strategies are being used. Additional research is needed to better understand these strategies and the types of examinees who adopt them.

Conclusions

The results of this study add to the evidence regarding the dynamics of SAT. The picture that is emerging, however, appears increasingly complex. Several distinct examinee variables have been found to moderate the effects of SAT, including test anxiety, verbal self-concept, and achievement locus of control. Moreover, this study is the first to find evidence of examinees for whom SAT has a negative effect.

Examinees can be affected by a variety of affective and motivational factors. Research on SAT has served to illustrate that these factors can exert a non-trivial influence on examinee test performance, even in situations where an IRT model can be shown to fit to a conventionally acceptable degree. If the relationships between these non-cognitive factors and test performance can be better understood, a new generation of computer-based tests may emerge that can adapt, not only to examinee ability, but to non-cognitive factors as well. Such

tests may be more effective in eliciting maximal examinee test performance than current computer-based tests.

References

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

Baron, R. S. & Logan, H. (1993). Desired control, felt control, and dental pain: Recent findings and remaining issues. *Motivation and Emotion, 17*, 181-204.

Burger, J. M. (1989). Negative reactions to increases in perceived personal control. *Journal of Personality and Social Psychology, 56*, 246-256.

Lefcourt, H. M. (1982). *Locus of control: Current trends in theory and research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lefcourt, H. M., Von Baeyer, C. L., Ware, E. E., & Cox, D. J. (1979). The Multidimensional-Multiattributional Causality Scale: The development of a goal specific locus of control scale. *Canadian Journal of Behavioral Science, 11*, 286-304.

Rocklin, T. (in press). Self-adapted testing. *Applied Measurement in Education*.

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

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

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

Sarason, I. G. (1986). *Test anxiety, worry, and cognitive interference*. In R. Schwarzer (Ed.), Self-related cognitions in anxiety and motivation. Hillsdale, NJ: Lawrence Erlbaum Associates.

Spielberger, C. D. (1980). *Preliminary professional manual for the Test Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

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

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

Vispoel, W. P., Rocklin, T., & Wang, T. (in press). Individual differences and test administration procedures: A comparison of fixed-item, computer-adaptive, and self-adaptive testing. *Applied Measurement in Education*.

Vispoel, W. P., Wang, T., de la Torre, R., Bleiler, T., & Dings, J. (1992, April). *How review options and administration modes influence scores on computerized vocabulary tests.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

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

Wise, S. L. (in press). Understanding self-adapted testing: The perceived control hypothesis. *Applied Measurement in Education*.

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

Table 1

Analysis of Variance for Estimated Proficiency

Source	SS	df	MS	F	F-Probability
Locus of Control (LOC)	.04	1	.04	.04	.842
Test Anxiety (TA)	18.20	1	18.20	17.54	<.001
Test Type (TT)	.65	1	.65	.62	.430
TT within Internal LOC	2.75	1	2.75	2.66	.104
TT within External LOC	6.45	1	6.45	6.23	.013
LOC x TA	.17	1	.17	.16	.685
LOC x TT	8.19	1	8.19	7.89	.005
TA x TT	.85	1	.85	.82	.365
LOC x TA x TT	.32	1	.32	.31	.581
Error	369.38	356	1.04		

Table 2

Descriptive Statistics for Estimated Proficiency by Test Type and Achievement Locus of Control

Achievement Locus of Control	Test Type					
	SAT			CAT		
	Mean	SD	n	Mean	SD	n
Internal	.28	1.09	173	.06	.97	163
External	-.52	1.26	12	.37	.93	22
All Examinees	.22	1.12	185	.10	.97	185

Table 3

Analysis of Variance for Posttest State Anxiety

Source	SS	df	MS	F	F-Probability
Locus of Control (LOC)	142.81	1	142.81	1.25	.265
Test Anxiety (TA)	6409.83	1	6409.83	55.91	<.001
Test Type (TT)	641.58	1	641.58	5.60	.019
TT within Internal LOC	1095.04	1	1095.04	9.55	.002
TT within External LOC	472.74	1	472.74	4.12	.043
LOC x TA	131.14	1	131.14	1.14	.286
LOC x TT	758.96	1	758.96	6.62	.010
TA x TT	236.57	1	236.57	2.06	.152
LOC x TA x TT	44.35	1	44.35	.39	.534
Error	40816.35	356	114.65		

Table 4

Descriptive Statistics for Posttest State Anxiety by Test Type and Achievement Locus of Control

Achievement Locus of Control	Test Type					
	SAT			CAT		
	Mean	SD	n	Mean	SD	n
Internal	35.95	11.49	171	40.07	10.98	161
External	47.42	12.45	12	39.95	14.71	21
All Examinees	36.70	11.87	183	40.05	11.42	182

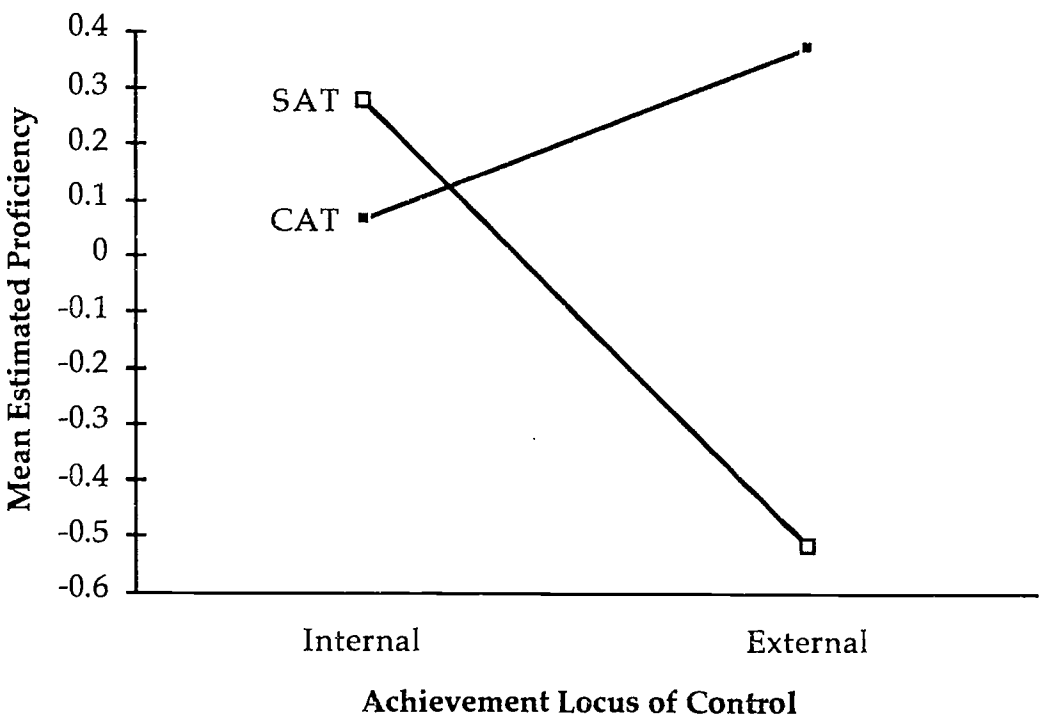

Figure 1. Interaction between test type and achievement locus of control for estimated proficiency.

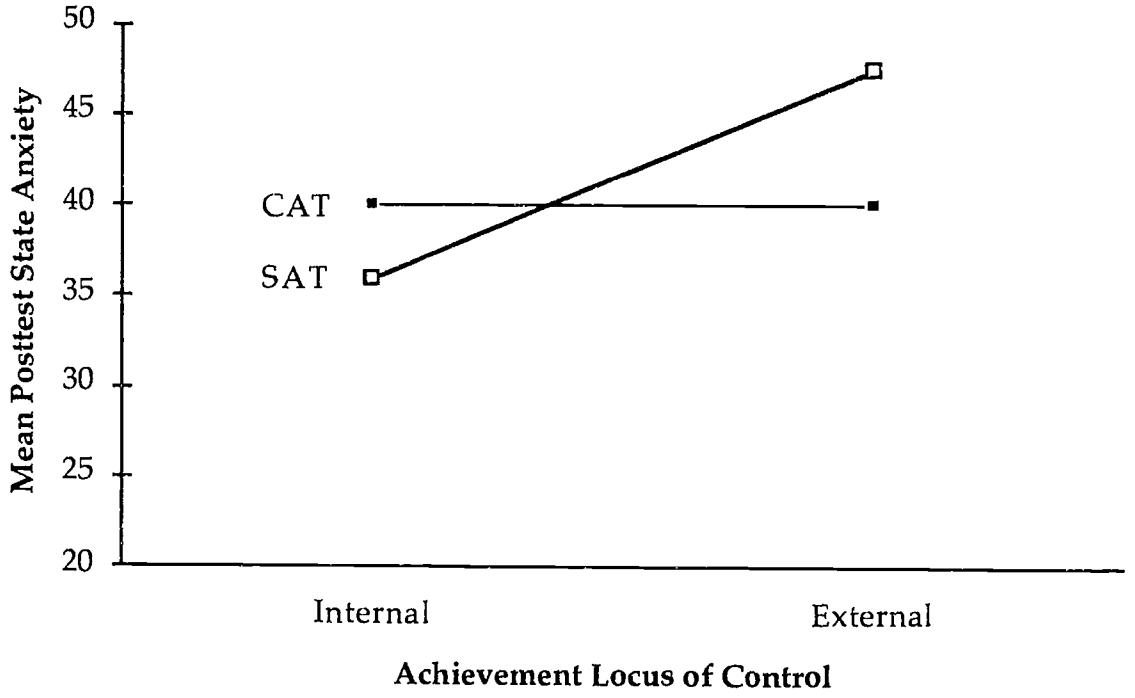

Figure 2. Interaction between test type and achievement locus of control for posttest state anxiety.