DOCUMENT RESUME

ED 383 725 TM 023 148

AUTHOR	Tobias, Sigmund; Everson, Howard
TITLE	Development and Validation of an Objective Measure of Metacognition.
PUB DATE	Apr 95
NOTE	36p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*College Students; Elementary School Students; Evaluation Methods; Grade 5; Higher Education; Intermediate Grades; Knowledge Level; *Mathematics Achievement; Mathematics Anxiety; *Metacognition; *Objective Tests; Research Methodology; Scoring; *Self Evaluation (Individuals); *Test Construction; Test Validity; Word Recognition
IDENTIFIERS	Monitoring

ABSTRACT

Four studies are reported that used a metacognitive evaluation procedure that can be group-administered and objectively scored. The procedure assesses the knowledge monitoring component of metacognition by evaluating the discrepancy between students' estimates of how well they are likely to perform on a task and their actual performance. The first study examined mathematics and mathematics anxiety in 51 fifth graders. Another study examined whether the metacognitive evaluation procedure was related to a more distant domain, such as learning in school. Participants were 139 college students (84 with complete data) taking a word knowledge test. The third study examined correlations of scores from the metacognitive evaluation procedure and prior learning in college for 115 students. The fourth study investigated the relationship between student estimation of words they would know and their estimates of performance on examinations for 77 college students. Results of the four studies confirm the importance of metacognitive monitoring on achievement in mathematics for elementary students and for college learning. Two tables and six figures present study findings. (Contains 43 references.) (SLD)

Development and Validation of an Objective Measure of Metacognition

Sigmund Tobias
City College of New York

Howard Everson
The College Board

A paper presented at a symposium entitled *Issues in Metacognitive Research & Assessment* at the annual meeting of the American Educational Research Association April 1995 San Francisco

Development and Validation of an Objective Measure of Metacognition

Sigmund Tobias Howard Everson
City College, CUNY The College Board

Metacognition can be defined as the ability to monitor, evaluate, and make plans for one's learning (Flavell, 1979; Brown 1980). Students with effective metacognitive skills have been found to be capable of activities such as the following: a) making accurate estimates of what they know and do not know, b) monitoring their on-going learning activities, and c) developing plans to learn new material. A large body of research, reviewed by Brown and Campione (1986) and by Baker (1989), has reported differences in metacognitive abilities between learning disabled and regular students, as well as between generally capable learners and their less able counterparts. This literature clearly indicates that metacognitive abilities are critically important for effective learning. It is, therefore, not surprising that metacognition is one of the most frequently studied constructs in contemporary cognitive instructional. and educational psychology.

The purpose of this paper is to report on four studies using a metacognitive evaluation procedure which may be group administered and objectively scored. The procedure assesses the knowledge monitoring component of metacognition by evaluating the discrepancy between students' estimates of how well they are likely to perform on a task and their actual performance. It is reasoned that the smaller the difference between estimated and actual performance, the better students' ability to monitor their knowledge and learning, one of the critical components of metacognition.

Assessing Metacognition

Despite its importance in meaningful human learning, the assessment of metacognition has proven to be both difficult and time consuming (O'Neil, 1991), creating a considerable obstacle to the advance of research in this field. Metacognition is usually assessed by inferences from performance, by ratings based on interviews in which students are questioned about their knowledge and processing strategies, and by analysis of "think-aloud" protocols (Meichenbaum, Burland, Gruson, & Cameron, 1985). Making such assessments for research purposes usually implies many of the following: students have to be examined singly, their learning actions observed closely, protocols of their cognitive activities have to be recorded and transcribed, content analyses of the protocols conducted and, finally, ratings of the protocols made in order to determine students' metacognition. As Royer, Cisero, and Carlo (1993) have observed "The process of collecting, scoring, and analyzing protocol data is extremely labor intensive" (p. 203), and the resources for such work are rarely available in most instructional situations, or in many university based research projects.

Labor intensive practices such as those described above make it difficult to evaluate metacognition in many instructionally relevant settings, including secondary and post-secondary schools, as well as training environments in business, industry, and governmental agencies. In view of these difficulties it is not surprising that metacognitive research is usually conducted in elementary school settings in which the time of participants can easily be diverted for the research effort. Our goal in the four studies reported here is to describe a technique for evaluating the ability to monitor one's knowledge and learning by assessing the discrepancy between students' estimates of how well they will perform a task and their actual performance. The measure can be group-administered and scored objectively. Furthermore, we also sought to relate the procedure to students' performance in the general domains of word knowledge and mathematics, two critical subjects in schooling at all levels. Finally, the studies described also examined the relationship of the procedure to students' college learning, and to test anxiety.

A number of self-report measures of metacognition (Everson, Hartman, Tobias, & Gourgey, 1991; O'Neil, 1991; Pintrich, Smith, Garcia, & McKeachie, 1991) have been developed. Such measures also have the advantage of being easily administered to groups and may be scored rapidly and objectively. Unfortunately, the use of self-report measures raises a variety of questions including some of the following: Are students aware of the cognitive processes used during learning? Are they able to describe and report on the processes used? Do they report them truthfully?

The metacognitive evaluation procedure used in the research to be reported assesses metacognition by evaluating the discrepancy between estimated and actual performance. In an earlier study (Tobias, Everson, Hartman, & Gourgey, 1991) the metacognitive evaluation procedure in the domain of word knowledge was administered to 167 college students. Some of the participants received a list of 33 words and were asked to check off the words they knew and did not know, and then responded to a vocabulary test based on the same words. The other groups read a two and a half page text passage in which all of the words on the vocabulary test were defined, before responding to the word list and vocabulary test. The procedure generated the following scores: student estimated that the word was a) known and answered correctly on vocabulary test [+ +], b) known and answered incorrectly on test [+ -], c) unknown yet was answered correctly on test [- +], d) unknown and answered incorrectly on test [- -]. Of course, the + + and - - scores represented accurate metacognitive estimates about students' vocabulary knowledge, while the others were indicative of inaccurate estimates.

The results indicated that accurate metacognitive judgments about the total number of words students thought they knew and actually knew had a substantial positive relationship with reading comprehension; similar estimates for the total number of words thought to be and actually unknown were negatively related to comprehension. All the metacognitive scores were significantly related to reading comprehension. However, the scores of students who read the text passage, in which the words were defined, before responding to the word list and vocabulary test had significantly higher relationships than the group who merely completed the word list and vocabulary test. The relationships between words estimated and actually known and reading comprehension for the latter group was .29, compared to .65 for the group who read the text passage. These findings were similar to the results reported by Everson, Smodlaka, and Tobias (1994) who administered only the word list and vocabulary test to study relationships with reading comprehension and anxiety. In that study a correlation of .35 was found between d' score, derived from signal detection theory (Green, & Swets, 1966; Macmillan, & Creelman, 1991), and reading comprehension.

Like other metacognitive measures, estimates of how well students are likely to perform on a test also rely on self-reports. However, it is reasoned that such reports should be more readily available to students than is their recollection of the types of cognitive processes engaged in during a preceding task, and/or how frequently the processes were used. The vocabulary test is not based on self-reports, but consists of students' actual test performance. Since estimated and actual performance can both be scored objectively, the procedure has a clear-cut advantage over asking students to report on their cognitive processes either in the form of protocols, or by self-report inventories.

In the four studies to be reported in this paper, the metacognitive evaluation procedure was extended to mathematics in the first investigation, which also examined its relationship with anxiety. The other studies examined the relationships of a revised version the word knowledge version of the metacognitive evaluation procedure to students' learning and estimates of test performance in college.

Study I: Assessing Metacognition in Mathematics, and Relationships with Math Anxiety[1]

Haneghan and Barker (1989) reported a number of investigations dealing with the effects of metacognition on the accuracy of problem representation. The studies indicated that metacognition was as important for the learning of mathematics as it was for reading. These results are supported by the expectations and findings of other researchers, such as Campione, Brown, and Connell (1989), Lester, Garofalo, and Kroll (1989), as well as Schonfeld (1985). Furthermore, research (Cardelle-Elawar, 1992; Montague, 1992) has also shown that students' performance in solving mathematical problems was facilitated when instructed with a metacognitive approach. Therefore, it was expected that the metacognitive evaluation procedure should be related to achievement in mathematics generally, and to students' ability to solve mathematical problems specifically.

A further purpose of this study was to investigate the impact of test anxiety on metacognition. It was suggested "that the relationship between test anxiety and metacognition may be a worthwhile field for research, while simultaneously helping to establish links between affect and cognition more generally" (Tobias, 1992, p. 28). High test anxiety has been found (Sarason, 1987) to lead students to divide their attention between the task and negative personal pre-occupations. It has been suggested (Tobias 1992, 1985) that interference in students' performance as a result of high anxiety was attributable to reduced cognitive capacity available for task solution. It was reasoned that the central representation of anxiety must absorb some proportion of cognitive capacity, leaving a reduced amount available for task solution. The further absorption of capacity required by metacognitive monitoring of cognitive processes was expected to be especially debilitating for highly anxious students whose cognitive capacity is expected to have been reduced by the central representation of test anxiety. Therefore, a negative relationship between anxiety and metacognition was expected since "highly test anxious students can be expected to have less adequate metacognitive abilities than those with lower anxiety" (Tobias, 1992, p. 28).

Method

A metacognitive evaluation procedure was developed for mathematics in which students were initially asked to estimate whether they could answer a group of mathematical questions correctly, and then asked to actually solve the problems. Measures of mathematical achievement, mathematical anxiety, and anxiety engaged by the task were also administered.

Procedures

The Fenema-Sherman (1976) scales assessing math anxiety and attitudes towards mathematics were administered in a first session. These were Likert type scales with five alternative responses. In order to assure that the participating elementary school students were able to read the questions, each item was read aloud to them.

A list of 30 mathematical questions was constructed made up of 20 items involving computation and 10 problem solving questions. The items were selected from the fifth grade mathematics curriculum. The math problems were administered during a second session, and students asked to determine if "you feel able solve these problems. Do not solve them now." Students were encouraged not to spend too much time on each item and asked to check one of two alternate spaces next to each problem, indicating whether they felt that they could, or could not solve the problem. This procedure was completed in six minutes, giving students an average of 12 seconds per problem to estimate whether they could solve each item.

During the third session, the same 30 questions were re-administered in the form of a test, i.e., the students were now asked to actually solve the problems. A total of 40 minutes was allocated to enable the students to complete the test. Immediately before and after working on the mathematical problems, the Worry-Emotionality (Morris, Davis, & Hutchings, 1981) scale, a 10 item Likert type measure of the degree to which anxiety was engaged at the moment, was administered. Students' mathematical achievement was determined from their scores on the Metropolitan Achievement Test (1985) obtained from the school files.

Participants. A total of 51 fifth grade students (31 females) from an urban public school served as subjects in this study. The students were predominantly of Hispanic origin, and their reading and mathematical achievement ranged from average for their grade, to two years below grade level.

Results and Discussion

The mathematics test was scored with reference to students' estimates of their ability to solve the problems. That procedure generated these four scores, for each student: (a) + +, felt that could solve the problem and did so, (d) - -, felt that could not solve problem and did not, (c) + - , felt that could solve problem, but did not, and (c) - +, felt that could not solve problem, but did.

Since there were no differences attributable to gender either on students' metacognitive estimates or their test anxiety, the data were pooled for further analysis. The four scores were correlated with the total math score on the Metropolitan Achievement Test (1985) obtained from the students' records. The correlations are displayed in Table 3. The "scores" column in that table represents the number correct

Insert Table 1 about here

on the math test. The + + and - -scores were combined to indicate correct estimates of students' ability to solve mathematical problems. Similarly the - + and + - scores were combined to form the incorrect estimates. The last two columns of Table 1 display the data dealing with those two variables. Table 1 also presents the correlations between the metacognitive scores and the scales assessing Mathematics Anxiety (scored in the direction of higher anxiety yielding higher scores) and Attitudes Towards Mathematics developed by Fenema and Sherman (1976), as well as with the Worry and Emotionality (Morris et al., 1981) components of test anxiety.

Table 1 indicates that each of the metacognitive estimates was significantly related to students' mathematics achievement. The correlation between number correct on the math test and Metropolitan score was .53. When that relationship is compared to the correlation of .73 between Metropolitan score and + +, or the correlation of .76 between the Metropolitan score and total number of correct estimates, it is clear that metacognitive estimates of the ability to answer the questions are more substantially related to mathematical achievement than the number of problems solved correctly, irrespective of estimate. That finding was confirmed by regression analysis. When the number of correct estimates, incorrect estimates, and total number right were in the model, only the correct estimates contributed significantly to the prediction of Metropolitan score (R Square Change = .08, $F(3,45)=8.52$, $p < .01$).

The finding that the metacognitive estimate accounts for variance in mathematics achievement above that attributable to number of problems solved correctly duplicates a similar finding in the Tobias et al.(1991)study in the domain of word knowledge where the ++ score had a correlation of .65 with reading comprehension, whereas the relationship between comprehension and total number of word correct was .45. The difference in the magnitude of these correlations in the present study, and the findings of Tobias et al. (1991) indicates that accurate metacognitive estimates contributed variance above and beyond the total number correct. In both investigations the scores based on estimated and actual performance accounted for about 4% more variance than the number correct alone. These results confirm the basic assumption of the metacognitive evaluation procedure that students' metacognitive judgments contribute significant independent variance beyond those accounted for by number correct on a test.

Table 1 also indicates that, as expected, relationships between the metacognitive evaluation procedure and mathematics anxiety were in the expected direction. Thus, mathematics anxiety was negatively related to incorrect estimates and positively related to correct ones The correlation between number right and the math anxiety score was -.25 and not significant, though the relationships with the metacognitive estimates were significantly related to anxiety. The negative relationships between metacognition and anxiety are generally similar to those found by Everson et al., (1994) who reported that anxious students had significantly fewer correct estimates than their less anxious counterparts, confirming expectations that anxious students have lower metacognition than their less anxious counterparts.

The results support predictions regarding the relationships between both achievement in mathematics and anxiety with the metacognitive evaluation procedure. As expected, there were significant and substantial correlations between students' metacognitive accuracy in estimating their ability to solve mathematical problems and their achievement in mathematics. Also as expected, inaccurate assessments were negatively related to achievement. While no causal inferences about mathematical achievement and metacognition can be made from these correlational data, the results indicate that the technique seems useful for further research in these areas.

A similar procedure for assessing metacognitive estimates in the mathematical domain was used in another study (Tobias, 1994), which also investigated the relationship of metacognitive estimates to participant's interests. The results of that investigation indicated that, as expected, the accuracy of students' metacognitive estimates of their ability to solve mathematical word problems increased from fourth through sixth grades. Also as expected, it was found that the accuracy of students' estimates increased with ratings of their mathematical ability. Those results, together with the findings of the present study, support the construct validity of the metacognitive evaluation procedure applied to the domain of mathematics.

Metacognition and College Learning

Prior research (Tobias et al.1991; Everson et al.1994) found that scores derived from the metacognitive evaluation procedure in the domain of word knowledge were significantly related to reading comprehension. Similarly, evidence of the metacognitive evaluation procedure's applicability to mathematics was seen from the results of Study I, reported in this paper, and from other findings (Tobias, 1994). The results of these investigations indicated that metacognitive estimates were closely related to competence in the domain in which students' estimates of knowledge were obtained, either reading comprehension or mathematical problem solving. One purpose of the next two studies was to examine whether the metacognitive evaluation procedure was related to a more distant domain than the one in which the assessment occurred, such as learning in school.

Another purpose of the succeeding studies was to extend the research on metacognition to students' learning in college. As indicated above, much of the research relating metacognition to school learning has been conducted in elementary schools, and to a lesser degree in and secondary school settings. The succeeding studies, to be described below, examined whether the metacognitive evaluation procedure in the domain of word knowledge was related to students' overall achievement in college, to their learning in different content areas.

It was reasoned that, in addition to assessing the discrepancy between estimated and actual performance, students' ability to estimate whether they have mastered new material is an important characteristic of effective learners at all educational levels, and especially in college. It was expected that those who could accurately estimate their word knowledge should be at an advantage in college settings, since they can use the available time to concentrate on what has not been mastered, and safely ignore what is already known. Students with less effective metacognitive knowledge monitoring abilities may waste time practicing or reviewing what they already know, rather than zeroing in on new material or updating partially learned content. Therefore, students' metacognitive accuracy in estimating their word knowledge was expected be related to their learning in college as reflected in their overall grade point average (GPA). Furthermore, in view of the importance of general word knowledge in English, humanities, and social and behavioral science courses it was expected that the highest relationships between metacognitive scores and GPA would be found in these classes compared to grades in science.

Materials

A revised version of the word knowledge materials used in prior research (Tobias et al., 1991; Everson et al., 1994) was employed in this investigation. In addition to some editorial revisions of the expository text used in one of the prior studies (Tobias et al., 1991) a narrative version of the same passage was also developed in order to examine the effect of situational interest on metacognition. The word list and vocabulary were also modified from those used previously, to contain an equal number of explicitly and implicitly defined words.

A total of 38 words were defined in the revised versions of the text, 19 words were explicitly defined (e.g., "Coronary or heart disease...."), and another 19 received implicit definitions (e.g., "Epidemiologists who have compared the prevalence of heart disease in the United States and in other countries..."). Explicit or implicit definitions were determined by two independent judges who rated all words. When there was any disagreement about a particular word, the judges conferred and the passage was modified to eliminate the disagreement. A multiple choice vocabulary test was developed, containing the correct choice and three distractors for the 38 items on the word list.

Procedures

The word list (alpha reliability= .99) and vocabulary test (alpha reliability= .80) were administered in a first session. The two versions of the text were then randomly assigned to students in a second session, followed by a re-administration of the word list and vocabulary test. The materials were administered to students during their classes in the presence of the instructors.

Participants

The sample consisted of 139 students attending a large urban university, though only 84 subjects completed all the materials during two sessions. Part of the sample consisted of students intending to obtain a college degree majoring in nursing. The nursing students ($N = 47$, $N = 33$ with complete data) were recruited from a class serving as the orientation course in a nursing program. The rest of the sample consisted of freshmen ($N = 92$, $N = 51$ with complete data) who were recruited from a freshman orientation course.

Results and Discussion

This report deals only with the relationship of the metacognitive evaluation procedure to students' college achievement; the results dealing with interest were reported elsewhere (Tobias, 1994, Study I). The correlation between total score on both administrations of the vocabulary test, based on 84 students who completed the test on both administrations, was .75. It should be noted that this is not a test re-test reliability coefficient since students read the text passage, containing explicit and implicit definitions of the words, immediately before the second administration of the vocabulary test.

Students' estimated word knowledge and performance on the vocabulary test were determined for both administrations. Two scores were computed for each administration: the total number of correct [words in the + + and -- categories] and incorrect estimates [+ - and - + categories]. Preliminary analysis indicated that there were no differences in these scores between students assigned to the expository and narrative versions of the text passage, so the data for both versions were pooled. The data for the ex- and implicitly defined words was also pooled. The correlations between the correct and wrong estimates on both administrations of the metacognitive evaluation procedure and students' overall GPA, and their grades in English, Humanities, Sciences and Social Sciences courses at the end of the term were computed and are shown in Table 2. Since 92 students were freshmen in their initial term of college the overall GPA

Insert Table 2 about here

for this group was based on an average of only 12.1 credits (SD=5.6), whereas the nursing students had a mean of 56.4 credits (SD=28.3). Therefore, the correlations are presented for each group separately, as well as for the total sample. The different number of cases in the various cells of Table 2 should also be noted.

In general correlations between the metacognitive evaluation procedure and GPAs shown in Table 2 provide encouraging evidence for the construct validity of the procedure. As expected, the correlations with the correct estimates are generally positive while those with the misses on both administrations are generally negative. The magnitude of the relationships between metacognitive scores and the overall GPA may have been affected by students taking courses in a variety of areas, including mathematics, physical education, and art, which are less likely to be related to metacognitive scores than subjects relying more heavily on reading. As expected, relationships with GPA in English and Humanities courses were generally higher than those in science, and overall GPA. Surprisingly, correlations between GPA in the social and behavioral sciences and the metacognitive scores were generally not significant. Perhaps grades in these courses, like science, reflect domain specific knowledge to a greater degree than the English and Humanities courses.

The significance of the correlations reported in Table 2 varies widely, probably as a function of at least three factors. First, the number of cases in each cell differs as a result of students' absence from either administration of the materials, leading to variability in the predictors. Second, students took a varying number of courses, and sometimes no courses at all, in some of the areas listed in Table 2, leading to variability in the criterion. Third, it is well known that college grades are often unreliable (Werts, Linn & Joreskog, 1978; Willingham, Lewis, Morgan, & Ramist, 1990). Furthermore, the reliability of the grades may have been reduced further by three factors: a) students took dissimilar courses, b) when similar courses were taken they were taught by different instructors, and c) by the differences in students' major fields of study. As expected, the correlations between metacognitive evaluation procedure scores and grades in English courses were generally higher, and more frequently significant, than those of any other subject. The findings indicate that metacognitive evaluation procedure scores are related to students' ability to learn materials from somewhat different domains than the ones on which the assessments were based. They also supported the concurrent validity of the procedure with respect to its relationship to learning in college.

For the 84 students with complete data for both administrations of the vocabulary test, the mean total score increased from 23.3 (SD= 6.0) for the first vocabulary test to 26.0 (SD=6.6) for the second ($t(83)= 5.53$, $p <.001$). Thus students clearly learned the meanings of some of the words after having the chance to update their word knowledge by reading the passage. However, the relationships between the metacognitive scores and grades shown in Table 2 were generally higher before students read the text passage, rather than afterwards, whereas the opposite results were reported in the Tobias et al. (1991) study. It should be noted that reading comprehension scores, not grades, were the criteria in the prior study, and since inferring the meaning of words is a key component of comprehension that may account for those findings in the Tobias et al. (1991) study. In the present investigation, it was assumed that having the chance to update one's word knowledge before estimating it would be more similar to students' learning in their classes than merely estimating word knowledge without any opportunity for new learning. Therefore, the relationships with grades were expected to be higher for scores from the second administration, after students had the chance to update their knowledge, than the first. The findings were generally not in accord with these expectations. The findings could not be attributed to students not learning very much from the text passage, since the metacognitive scores improved significantly from one administration to the next. It remains for further research to explore the reasons for these findings.

Evaluating the accuracy of students' word knowledge estimates before and after reading the text passage could be considered to be similar to dynamic assessment approaches (see Carson & Wiedl, 1992; Guthke, 1992; Lidz, 1992) in which students have the opportunity for new learning before being tested. Dynamic assessment procedures usually include some intervention in students' attempts to learn, observations of their reaction to the intervention, and an evaluation of students' responses to the assistance received as part of the assessment. Reviews have suggested (Carson & Wiedl, 1992) that students' attempts to verbalize learning difficulties and their receiving elaborated feedback contributes heavily to the value of dynamic assessment. The metacognitive evaluation procedure described here does not include any of these additional components typical of dynamic assessment, and is best considered to be a "test- opportunity to learn-retest" technique. A more active intervention, between the first and second administration of the word list and vocabulary test, designed to help students learn words from the passage, would have increased the similarity of the metacognitive evaluation procedure both to dynamic assessment approaches and to students' learning from their courses. It remains for further research to examine whether such a change will lead to higher relationships between students' metacognitive estimates after having the opportunity to learn the meanings of some words with their grades in college course.

Study III

The second study examined the correlations of scores from the metacognitive evaluation procedure and students' prior learning in college. The purpose of the third study was to investigate whether the procedure could be used to predict how well entering students would perform academically during their first year of college.

Procedure

The materials used in this study were identical to those employed in the second investigation. The metacognitive evaluation procedure was administered to the participants while they were attending a pre-freshman skills program prior to beginning their first semester of college. Students' achievement, in terms of GPA, was determined from the college records at the end of their first year in school.

Participants

The sample consisted of 115 students (59 Female) taking part in a pre-freshman skills program attended by students admitted to a large urban University. The program was conducted during the summer preceding students' start of college classes.

Results and Discussion

The word list and vocabulary data were scored to determine the number of correct metacognitive estimates of students' word knowledge. Correct estimates were defined in the same way as in Study II. Again, preliminary analysis indicated that there were no differences between the expository and narrative passages, nor between the words defined explicitly or implicitly. Therefore, these data were pooled in the succeeding analyses.

The grades received by the students after their first year of college in the following content areas were obtained: English, Humanities, Social Science and Science courses. These grades were combined into an overall content area GPA. In Study II, correlation analysis was the optimal mode for analyzing the data due to the large number of students who were absent from the second administration of the materials, and the varying number of credits taken by the two groups of students. Correlations, by examining whether increases in one variable were accompanied by increases in the other, were also likely to maximize errors attributable the low reliability of grades (Werts et al., 1978; Willingham et al., 1990). Since all the participants in this study were incoming freshmen, the number and types of courses taken by students were more similar than in the earlier investigation. Therefore, the data were analyzed by creating high and low achievement groups by splitting students at the median of the GPA distribution on each of a number of academic areas, and on the combined GPA. Analysis of variance was then computed to determine the differences between students above and below the median GPA to determine the significance of differences on their metacognitive scores. Furthermore, the differences between the first and second administrations of the metacognitive evaluation procedure were analyzed as a within subjects second level of the ANOVA.

As expected, the results indicated that students above the median GPA made significantly more accurate overall metacognitive judgments, $F(1,113)=6.51$, $p < .05$, than those below the median. Also as expected, there was a significant difference between the metacognitive scores students obtained on the first and second administrations ($F(1,113)=15.19$, $p < .01$) of the word list and vocabulary test, thought there was no interaction between these variables. These data are displayed in Figure 1.

<u>Insert Figure 1 about here</u>

High and low groups in English, Humanities, Science, and Social Science courses were also formed by splitting the students at the GPA median in each of these content areas and examining the significance of differences on the number of correct metacognitive estimates. In English the overall metacognitive differences between students above and below the median in that subject were significant($F(1,113=5.62$, $p=.02$), as were the differences between the first and second administrations ($F(1,113)=89.29$; $p<.001$); there was no interaction. These data are shown in Figure 2.

<u>Insert Figure 2 about here</u>

The overall differences in metacognitive accuracy for students above and below the median in Humanities courses (Art, History, Music, Philosophy, World Civilization, World Humanities, and World Arts) were also significant ($F(1, 113) = 8.06$, $p < .01$), as were the differences between first and second administrations ($F(1, 113) = 13.58$, $p < .001$), and again there was no interaction. These data are shown in Figure 3.

<u>Insert Figure 3 about here</u>

Metacognitive differences between those above and below median GPA in the Sciences and Social Sciences were not significant.

The relationships between metacognitive scores and GPA at the end of the first year of college were generally similar to those reported in Study II. These results provide encouraging support for the importance of metacognition in predicting learning in a domain somewhat different from that in which the construct was assessed, and for the usefulness of the metacognitive evaluation procedure. Many of the participants had been recommended for the pre-freshmen program because they were considered to be at risk for poor performance in college. It seems likely that this factor reduced the range of college achievement for the sample and, therefore, may also have reduced metacognitive differences between the groups. Even though data were not collected in sections of the pre-freshmen skills program devoted exclusively to English as a Second Language (ESL), some of the students were signed up for both ESL and other skills sections, and thus ended up as part of the sample. The presence of non-native English speakers could also have reduced group differences in this study. Further research limited to native English speakers who were somewhat more heterogeneous with respect to possession of academic skills is needed to determine whether metacognitive differences between low and high achieving students are greater than those found in this study.

Another factor limiting the reliability of the findings was the fact that many of the students in the present sample took less than a full-time schedule of courses. Therefore, in order to increase the reliability of the criterion, it would also be useful to investigate the predictive validity of the metacognitive evaluation procedure in settings with a greater percentage of full-time students.

Study IV

Studies II and III examined the relationships between the metacognitive scores and students' grades in college. The grades received by students are a function not only of their domain knowledge, but also of the types of evaluations administered by instructors, as well as their grading practices. These latter factors potentially add some error into the relationship between metacognitive scores and GPA. In view of the fact that the metacognitive evaluation procedure assesses students' ability to estimate their knowledge, it should also be related to students' estimates of their performance on examinations. It was reasoned that students who were capable of accurately estimating the words they know and do not know, should also be more accurate in predicting how well they will perform on examinations based on content related to their present studies before they take them, and how well they performed on those examinations after they were completed. The fourth study tested these expectations.

There has been some research o[illegible]ediction of performance in courses and tests, though none of the studies has related the predictions to metacognition. Keefer (1971) found that college students who accurately estimated their performance achieved at a significantly higher level than less accurate estimators, and had a more positive self concept than their low estimating counterparts. Holen and Newhouse (1976) found that students' predictions of their grades on a course examination correlated as highly with actual performance as their GPA, and were significantly more accurate predictors than other variables such as grades in pre-requisite courses, or high GPA. Furthermore, students' predictions, contributed significant unique variance to predictions of actual final grade, above that contributed by high school and college GPA, or grades in prerequisite courses. Harris (1990) found that accurate estimators of test performance in psychology earned a significantly higher final average in introductory psychology than did low and less accurate estimators. Since Study II and III found that accurate metacognitive assessments were associated with higher GPA, the findings dealing with performance estimation support the rationale that students who make accurate metacognitive assessment of their knowledge should make more accurate predictions of test scores than less accurate students.

In Studies II and III all students responded to the metacognitive procedure before and after reading the text passage. It will be recalled that the results of Study II indicated that metacognitive estimates before students read the text passage were somewhat more highly related to their GPAs than those obtained after reading the passage. A further purpose of the fourth study was to vary the administration of the text passage, in order to examine its contribution to students' estimates of their test performance.

Method

The Advanced Placement (AP) Examination in Psychology (College Board, 1992) was administered to students enrolled in an introductory psychology class, who were asked to predict how many items they were likely to get right on that examination before it was taken, and after it was completed.

Procedures

Half the sample (n=39) was randomly assigned to read the expository version of the text passage used in the two preceding studies before they responded to the word list, while the other half (n=38) received an irrelevant task, the text selection titled "Teaching the Mentally Retarded" from Royer's Sentence Verification procedure (Royer, Carlo, Dufresne, & Mestre, 1994), and then answered the questions on that passage. The same word list and vocabulary test used in Studies II and III were then administered to all participants.

Students were then given a description of the different areas covered by the 1992 form of the AP Psychology test (College Board, 1992) and asked to predict how many of the 100 items they would answer correctly on that test. After completing the AP, students were asked to estimate how many items they had answered correctly.

Participants

A total of 77 students (41 females) taking the Introductory Psychology class on one of the campuses of a large urban university volunteered to participate in the study. Students could choose from a number of projects to satisfy a requirement for participating as subjects in research.

Results and Discussion

More accurate metacognitive scores were expected for the group responding to the word list and vocabulary test after reading the text compared to the other group who received the SVT, which was irrelevant to the task. Surprisingly MANOVA based on the total number of accurate estimate [+ + and - -] scores revealed no significant differences between the groups. Examination of the basic eight scores [+ +, + -, - +, - -, for both ex- and implicitly defined words] indicated that there appeared to be group some group differences, see Figure 4, but that these were reduced when the data were combined into total number of correct estimates, see Figure 5.

Insert Figures 4 & 5 about here

When MANOVA was computed on six of the basic scores (the scores for the + - category for ex-and implicitly defined words were eliminated to avoid linear dependencies) the differences between the groups were significant (Wilks Lambda= .76, Approximate $F(6,70)=3.71$, $p<.01$). Univariate F tests indicated that the students made more accurate metacognitive estimates on explicitly defined words in the + + category ($F(1,75)=5.97$, $p<.02$), and had fewer explicitly defined words in the - - category, ($F(1,75)= 4.74$, $p<.05$).

It was expected that students high on metacognitive knowledge monitoring would be more accurate in estimating their actual and estimated scores on the AP test before and after completing it, as well as obtain higher scores on the test. Finally, as suggested by other studies of student's estimation of their performance, it was expected that they would expect to obtain higher scores in the course in which they were registered. These predictions were tested by splitting students at the median on total number of accurate metacognitive predictions [combining the + + and - -] and computing MANOVA to examine the significance of the differences on students' estimates of their AP scores before and after taking the test, their actual AP score, and their expected final grade in the psychology class. No differences between groups who did, and did not read the text passage were expected or obtained in estimates regarding the AP test (Wilks Lambda= .980, Approximate $F(4,69) < 1$). The results for high and low metacognitive groups indicated that there was an overall difference between the groups (Wilks Lambda= .859, Approximate $F(4,69)= 2.83$, $p<.05$). Univariate tests indicated that students in the high metacognition group obtained higher AP scores ($F(1,72)=7.81$ $p<.01$), and that differences in expected final grade in the course just failed of significance ($F(1,72)=3.41$, $p<.10$). The means for the expected grade and AP data for high and low metacognitive groups are shown in Figure 5. There was no interaction between group and metacognition.

Insert Figure 6 about here

In general the results of this study supported the construct validity of the metacognitive evaluation procedure. Students high in the ability to monitor their word knowledge, also obtained higher scores on the AP exam and expected higher grades in the course for which they were registered. The absence of group differences on predicted AP score before taking the test was not surprising since students were completely unfamiliar with the test, beyond being informed about the categories of knowledge covered. They had no information about the difficulty of the items, or the types of preparation expected for the test, or specifically what they would be questioned on. The absence of differences on students' score estimates after they had taken the test was a little more surprising, since participants now had a much clearer idea about what the test covered. Perhaps this brief exposure to the test was inadequate to familiarize them with the domain covered by the AP.

Ideally, of course, participants' performance estimates about both predicted and actual grades should have been studied in the course for which they were registered. In that case students have enough information to make more reasonable predictions based on their experience with the subject matter, instructors, and context of the courses. Unfortunately the rules in place on this campus regarding participation of human subjects, made it impossible to compare students' estimated and actual final grade in the course. Data have been collected, though not yet analyzed, on another campus to make such a comparison, and also to compare test grades for high and low metacognitive students.

General Discussion

The results of the four studies confirm the importance of metacognitive knowledge monitoring on achievement in mathematics for elementary school students, and for college learning. The findings indicate that the technique used to determine students' metacognitive word knowledge seems useful for future investigations of metacognition. As suggested previously (Royer et al., 1993, Tobias et al., 1991), the usual ways of evaluating metacognition by interview, protocol analysis, or by making inferences from students' performance are very labor intensive. That makes it difficult to assess the construct in many large-scale research situations involving administration of the materials and procedures to relatively large groups and, for convenience and economy, scoring students' responses rapidly or objectively. The metacognitive evaluation procedure used in this study avoids these difficulties, making it a useful alternative to the more traditional modes of assessing this construct.

It should be noted that the findings regarding metacognition should not be generalized to such metacognitive activities as planning or use of strategies, since these were not assessed in this study. However, the ability to differentiate between the known and the unknown, which is assessed by the metacognitive evaluation procedure, may be the most fundamental component of metacognition. It would be difficult for students to make reasonable plans for learning, or to select appropriate strategies if they can not differentiate between what they already know and what they still have to learn. It may be useful to investigate empirically whether relatively accurate monitoring of what is known and unknown is a prerequisite for effective planning and selection of strategies for succeeding learning.

The evaluation procedure yields a variety of scores by which metacognition can be assessed. The number of correct estimates (combining + + and - - scores) was found to be most appropriate in the first three studies, but not in the fourth. Clearly, further research is needed to determine which of the scores gives the best assessment of accurate knowledge monitoring. Ideally, procedures such as the analysis of covariance matrices should be used to examine which of the indices assess the latent variable, metacognition, most effectively. Such research, using larger samples than those used in most of these studies, is presently planned.

Relationship to Metamemory Research

The metacognitive evaluation procedure described in this paper is similar to metamemory research on the feeling of knowing (FOK) and judgment of learning (JOL). FOK judgments "occur during or after acquisition and are judgments about whether a given currently non-recallable item is known and/or will be remembered on a subsequent retention test.....Judgments of learning (JOL) occur during or after acquisition and are predictors about future test performance on currently recallable items" (Nelson & Nahrens, 1990, p. 130). In terms of that definition, students' judgments on both the word list and math problems in the preceding research were similar to JOLs.

FOK research was originated by Hart (1965) who asked general information questions and students, after failing to recall an item, were required to make a judgment regarding their FOK that item. Finally, they were asked to select an answer from a subsequent set of distractors. The procedure has been extended to asking students to guess whether they could recall words learned in a paired associate task (Hart, 1967; Ryan, Petty, & Wentzlaff, 1982). Nelson, Gerler, and

Nahrens (1984) also extended the FOK research to students' ability to relearn, and to perceptual identification tasks, and Reder and Ritter (1992) investigated whether students opted either to retrieve or calculate mathematical problems, and the latency and accuracy of these processes. A review of FOK research indicated that "a large number of studies confirmed that (students).... unable to retrieve a solicited item from memory can estimate with above chance success whether they will be able to recall it in the future, produce it in response to clues, or identify it among distractors....The standard finding is that the predictive validity of FOK judgments is above chance, though far from perfect" (Koriat, 1993, p. 609-610). The findings of the present studies certainly confirm the trend observed in the metamemory research.

The FOK and JOL paradigms differ from the present research in a number of ways. First, the FOK judgments are typically required after a recall failure, rather than after every stimulus presentation. Second, attempts are usually not made, in either FOK or JOL research, to enable students to learn and/or correct their knowledge of the stimuli, as they were in the present research. Third, the purposes of the metamemory research are to clarify the mechanisms accounting for FOK and JOL, rather than to use it as a measure of metacognition to be related to students' learning of school tasks, or to their interests (Tobias, 1994).

References

Baker, L. (1989). Metacognition, comprehension monitoring, and the adult reader. Educational Psychology Review, 1, 3-38.

Breland, H., Jones, R.J., & Jenkins, L. (1994). The College Board vocabulary study. College Board Report No. 94-4. NY: The College Board.

Brown, A.L. (1980). Metacognitive development and reading. In R.J. Spiro, B.B. Bruce, & W.F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 453-481). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Brown, A. L. & Campione, J.C. (1986). Psychological theory and the study of learning disabilities. American Psychologist, 14, 1059-1068.

Campione, J.C., Brown, A.L., & Connell, M.L. (1989). Metacognition: On the importance of understanding what you are doing. In R. Charles & E. Silver (Eds.) The teaching and assessing of mathematical problem solving (pp. 93-114). Reston VA: National Council of Teachers of Mathematics.

Cardell-Elawar, M. (1992). Effects of teaching metacognitive skills to students with low mathematics ability. Teaching and Teacher Education, 8, 109-121.

Carlson,-J,-S. & Wiedl,-K.-H. (1992). Principles of dynamic assessment: The application of a specific model. Learning-and-Individual-Differences; 4, 153-166.

College Board

Everson, H.T., Hartman, H., Tobias, S., & Gourgey, A. (1991, June). A metacognitive reading strategies scale: Preliminary validation evidence. Paper presented at the annual convention of the American Psychological Society, Washington, D.C.

Everson, H. T., Smodlaka, I., & Tobias, S. (1994). Exploring the relationsip of test anxiety and metacognition on reading test performance: A cognitive analysis Anxiety, Stress, and Coping, 7, 85-96.

Fenema, E. & Sherman, J.A. (1976). Fenemana-Sherman mathematics attitude scales. Journal for Research in Mathematics Education, 7, 324-326.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. American Psychologist, 34, 906-911.

Green, D.M., & Swets, J.A. (1966). Signal detection theory and psychophysics. NY: Wiley.

Guthke,-J. (1992). Learning tests: The concept, main researc'ch findings, problems and trends. Learning-and-Individual-Differences; 4, 137-151

Haneghan, J.P., & Baker, L. (1989). Cognitive monitoring in mathematics. In C.B. McCormick, G. Miller, & M. Pressley (Eds.), Cognitive strategy research: From basic research to educational applications (pp 215-238.) New York: Springer.

Harris, L.J. (1991) Performance estimation and achievement in introductory psychology. Perceptual and Motor Skills, 71, 1072-1074.

Hart, J.T., (1967). Memory and the memory monitoring process. Journal of Verbal Learning and Verbal Behavior, 6, 685-691.

Hart, J.T. (1965). Memory and the feeling-of-knowing experience. Journal of Educational Psychology, 56, 208-216.

Holen, M.C., & Newhouse, R.C. (1976). Student self-prediction of academic achievment. Journal of Educational Research, 69, 219-220.

Kiefer, K.E. (1971). Characateristics of students who make accurate and inaccurate self-predictions of college achievement. Journal of Educational Research, 64, 401-404.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. Psychological Review, 100, 609-639.

Lester, F.K., Garofalo, J., & Kroll, D.L. (1989). The role of metacognition in mathematical problem solving. Final report to the National Science Foundation. Grand number MDR 85-50346. School of Education, Indiana University, Bloomington.

Lidz,-C.(1992) Dynamic assessment: Some thoughts on the model, the medium, and the message. Learning-and-Individual-Differences, 4, 125-136

Macmillan, N.A., & Creelman, C.D. (1991). Detection theory: A user's guide. NY: Cambridge University Press

Meichenbaum, D., Burland, S., Gruson, L. & Cameron, R. (1985). Metacognitive assessment. In S.R. Yussen (Ed.), The growth of reflection in children (pp. 3-27). New York: Academic Press.

Metropolitan Achievement Test (1985). San Antonio TX: The Psychological Corporation.

Montague, M. (1992). The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. Journal of Learning Disabilities, 25, 230-248.

Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. Journal of Educational Psychology, 73, 541-555.

Nelson, T.O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgements for predicting perceptual identification and relearning. Journal of Experimental Psycholgy: General, 113, 282-300.

Nelson, T.O., & Nahrens, L. Metamemory: A theoretical framework and new findings. In G.H. Bower (Ed.), The psychology of learning and motivation (pp. 125-173). New York: Academic.

O'Neil, H.F., Jr. (1991). Metacognition: Teaching and Measurement. Paper delivered at the annual convention of the American Psychological Association, San Francisco.

Pintrich, P. R., Smith, D.A., Garcia, T., & McKeachie, W.J. (1991). A manual for the use of the Motivated Strategies for learning questionnaire (MSLQ). Ann Arbor, MI: National Center for Research to Improve Postsecondary teaching and learning.

Reder, L.M., & Ritter, F.E. (1992) What determines initial feeling of knowing? Familiarity with questions terms, not with answers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 435-451.

Royer, J.M., Carlo, M.S., Dufresne, R., & Mestre, J. (1994). The asessment of levels of dodmain expertise while reading. Unpublished manuscript, University of Massachusetts, Amherst, MA.

Royer, J.M., & Cisero, C.A., & Carlo, M. (1993). Techniques and procedures for assessing cognitive skills. Review of Educational Research, 63, 201-243.

Sarason, I. G. (1987). Test anxiety, cognitive interference, and performance. In R. E. Snow & M. J. Farr (Eds.), Aptitude, learning, and instruction (Vol 3): Conative and affective process analyses (pp.131-142). Hillsdale, NJ: Erlbaum.

Schonfeld, A.H. (1985). Mathematical problem solving. Orlando FL: Academic Press.

Tobias, S. (1994). Interest and metacognition in word knowledge and mathematics. Paper presented at the annual convention of the American Educational Research Association, New Orleans, LA., April.

Tobias, S. (1992). The impact of test anxiety on cognition in school learning. In K. Hagtvet (Ed.), Advances in Test Anxiety Research (Vol. 7, pp. 18-31). Lisse, Netherlands: Swets & Zeitlinger.

Tobias, S. (1985). Test anxiety: Interference, defective skills, and cognitive capacity. Educational Psychologist, 20, 135-142.

Tobias, S., Hartman, H., Everson, H., & Gourgey, A. (1991, August). The development of a group administered, objectively scored metacognitive evaluation procedure. Paper presented at the annual convention of the American Psychological Association, San Francisco, CA.

Werts, C., Linn, R.L., & Joreskog, K.G. (1978). Reliability of college grades from longitudinal data. Educational and Psychological Measurement, 38, 89-95.

Willingham, W.W., Lewis, C., Morgan, R., & Ramist, L. (1990) Predicting college grades: An analysis of institutional trends over two decades. NY: The College Board.

Footnotes

1) These data were collected by Dhalma Rosado.

Table 1. Correlations Betweeen Metacognitive Evaluation Procedure Scores and Selected Variables.

Variable	+ +	+ -	- +	- -	Score	Correct Estimates	Incorrect Estimates
Metropolitan[1]	.73**	-.43**	-.65**	-.11	.53**	.76**	-.72**
Math Anxiety[2]	-.42**	.32*	.38**	.00	-. 29*	-.46**	.46**
Math Attitude[2]	.28*	-.16	-.35**	.07	.12	.34**	-.32
Worry- Pre	-.13	.42**	.12	.13	-.09	-.09	.40**
Worry -Post	-.05	.04	.10	.06	-.03	.01	.09
Emotion-Pre[3]	-.33**	.05	.32**	.28*	-.22*	-.23*	.22
Emotion-Post[3]	-.19	.04	.25*	.14	-.08	-.14	.17

1= Metropolitan Achievement Test, Mathematics Battery.

2 = Fenema-Sherman

3= Emotionality Scale.

*= $p < .05$

**= $p < .01$

Table 2. Correlations Between Metacognitive Evaluation Procedure Scores and Overall Grade Point Averages and Grades in Different Subject Areas.

Variables	First Administration Estimates			Second Administration Estimates		
Group	Correct		Wrong	Correct		Wrong
Total GPA	n	r	r	n	r	r
Total	101	.20*	-.19*	94	.09	-.16
Freshmen	65	.09	-.08	61	-.10	.00
Nurses	36	.28*	-.28*	33	.19	-.15
English GPA						
Total	72	.30**	-.27**	63	.24*	-.22*
Freshmen	53	.31**	-.28*	48	.00	-.13
Nurses	19	.25	-.21	19	.45*	-.30
Humanities GPA						
Total	82	.26**	-.25**	74	.13	-.14
Freshmen	52	.12	-.13	46	-.11	.11
Nurses	30	.47**	-.44**	28	.35*	-.45**
Science GPA						
Total	65	.18	-.20	63	.03	-.26*
Freshmen	28	.11	-.12	27	-.28	-.15
Nurses	37	.26	-.29*	36	.18	-.45**
Social Science GPA						
Total	64	.18	-.20	63	.24*	-.40**
Freshmen	26	.15	-.20	29	.14	-.35*
Nurses	38	.09	-11	34	.14	-.29*

*= $p < .05$
**= $p < .01$

Metacognitive Hits by Content Area GPA.

Hi vs Low, $F=6.51$, $p < .05$

Pre vs Post $F=15.19$, $p < .001$

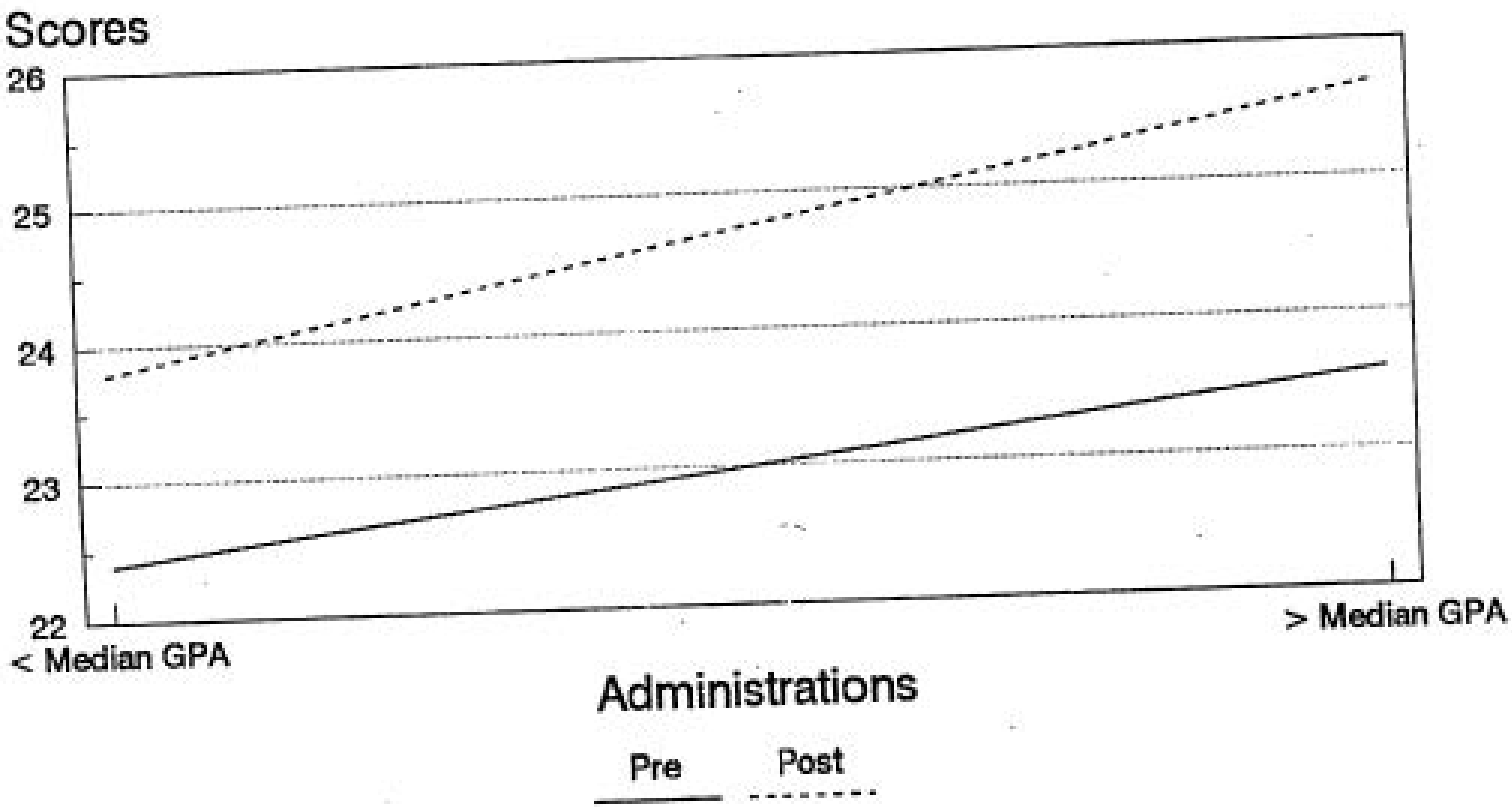


Figure 1

Metacognitive Hits by English GPA

High vs. Low, $F=5.62$, $p=.02$.

Pre vs. Post, $F=8.09$, $p<.01$

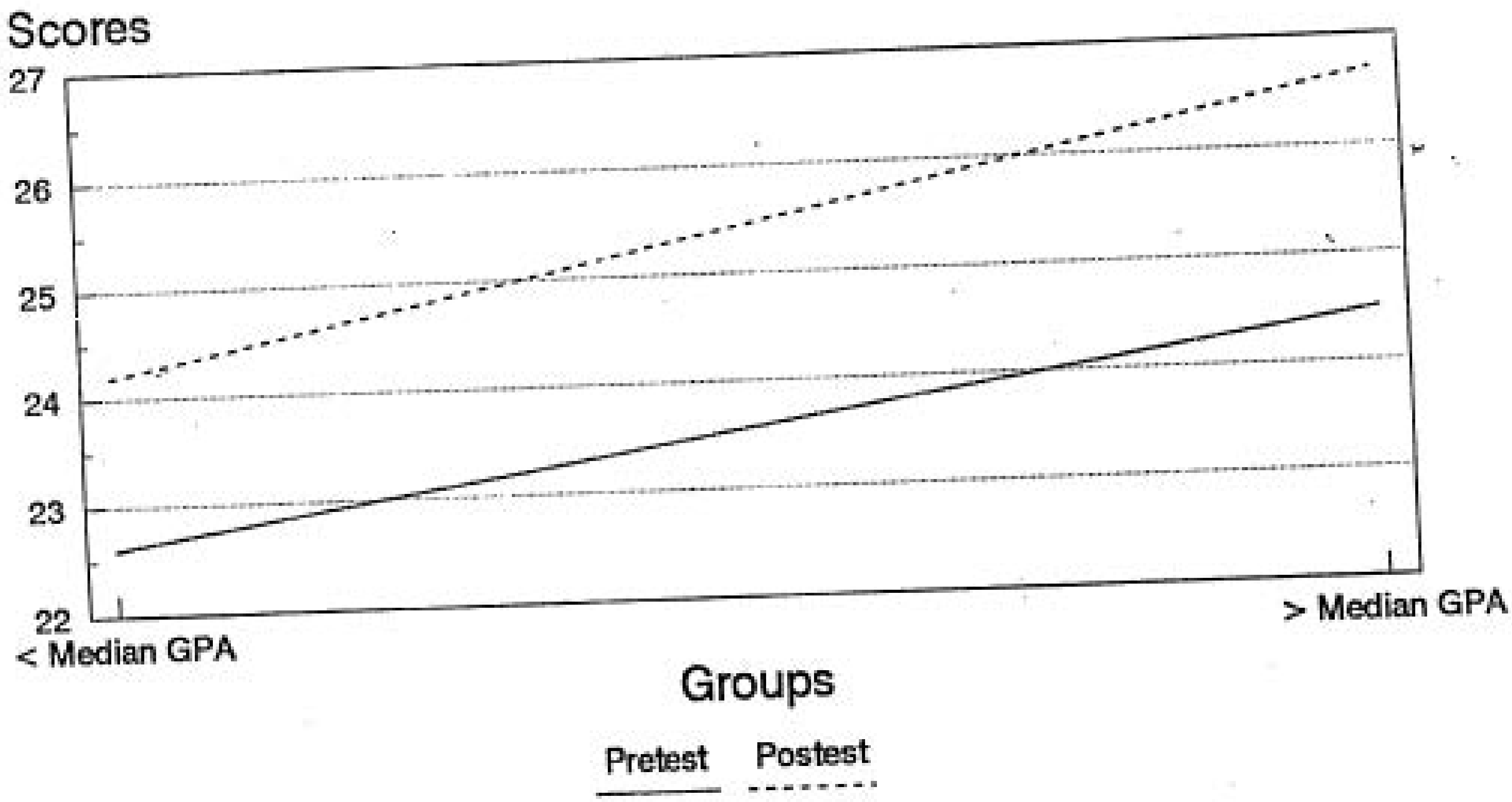
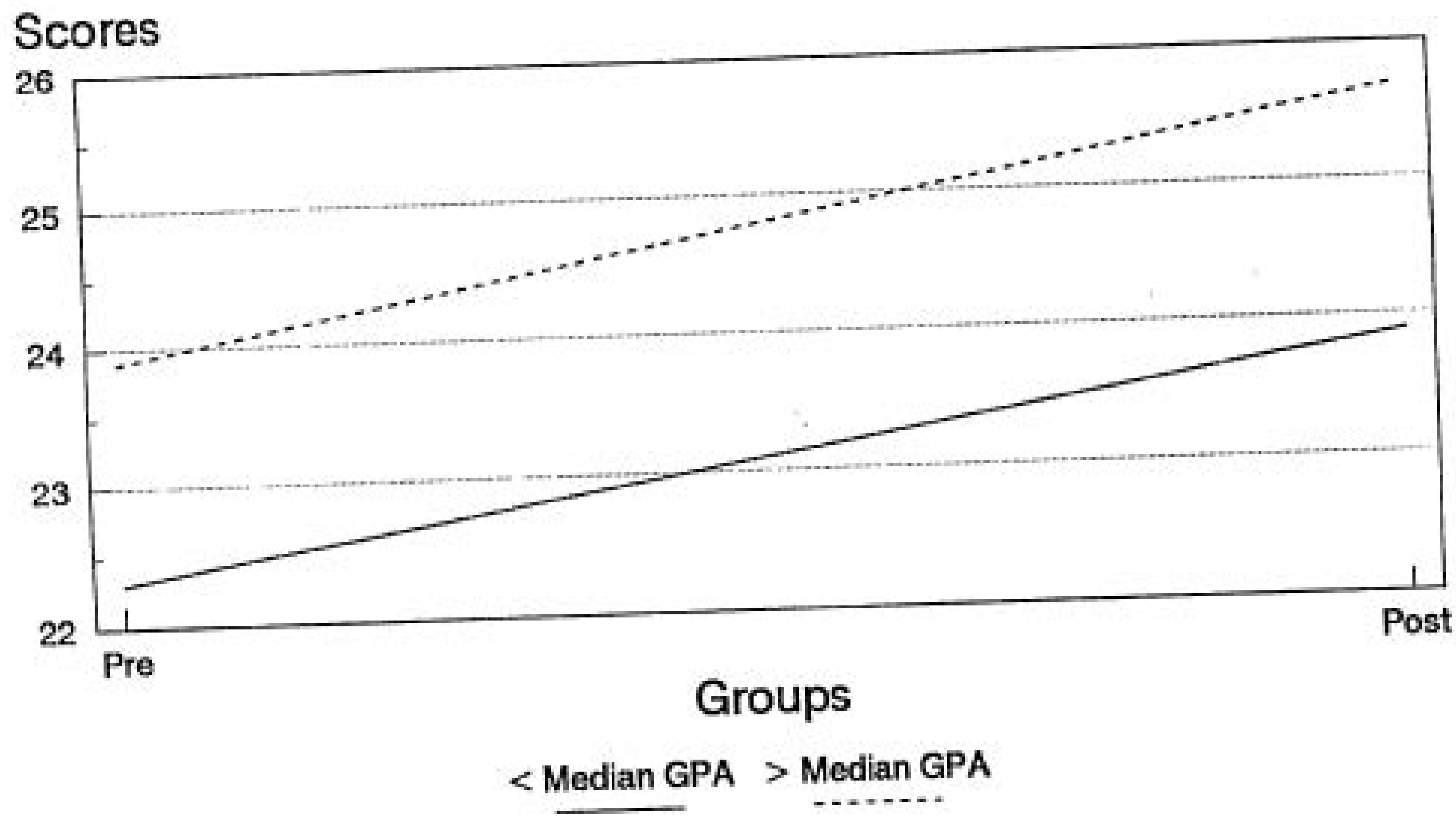


Figure 2

Metacognitve Hits by Humanities GPA

High vs Low GPA, F=8.09, p< .01

Pre vs. Post, F=13.58, p < .001



29

Figure 3

Basic Metacognitive Scores

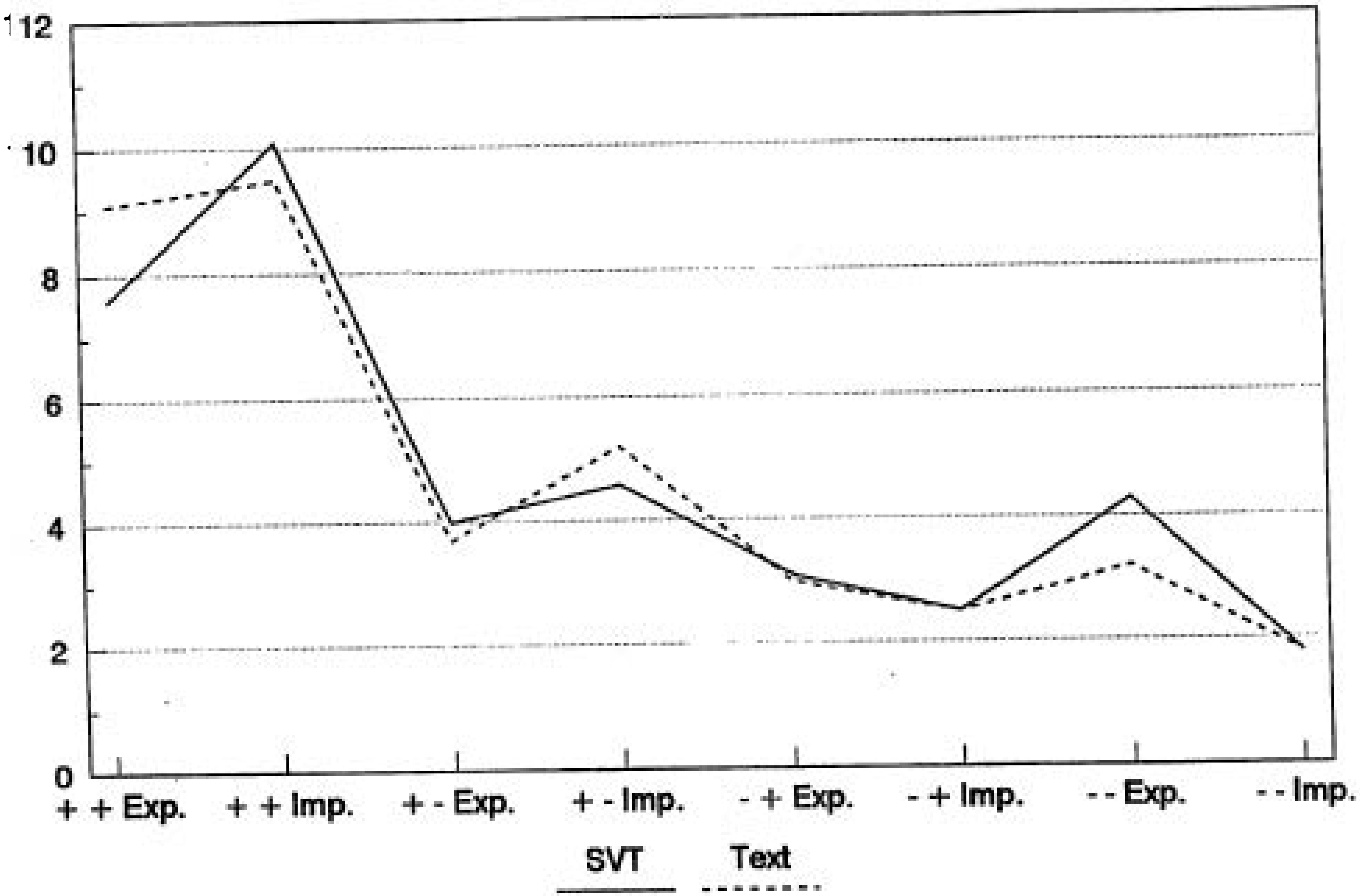


Figure 4.

Compound Metacognitive Scores

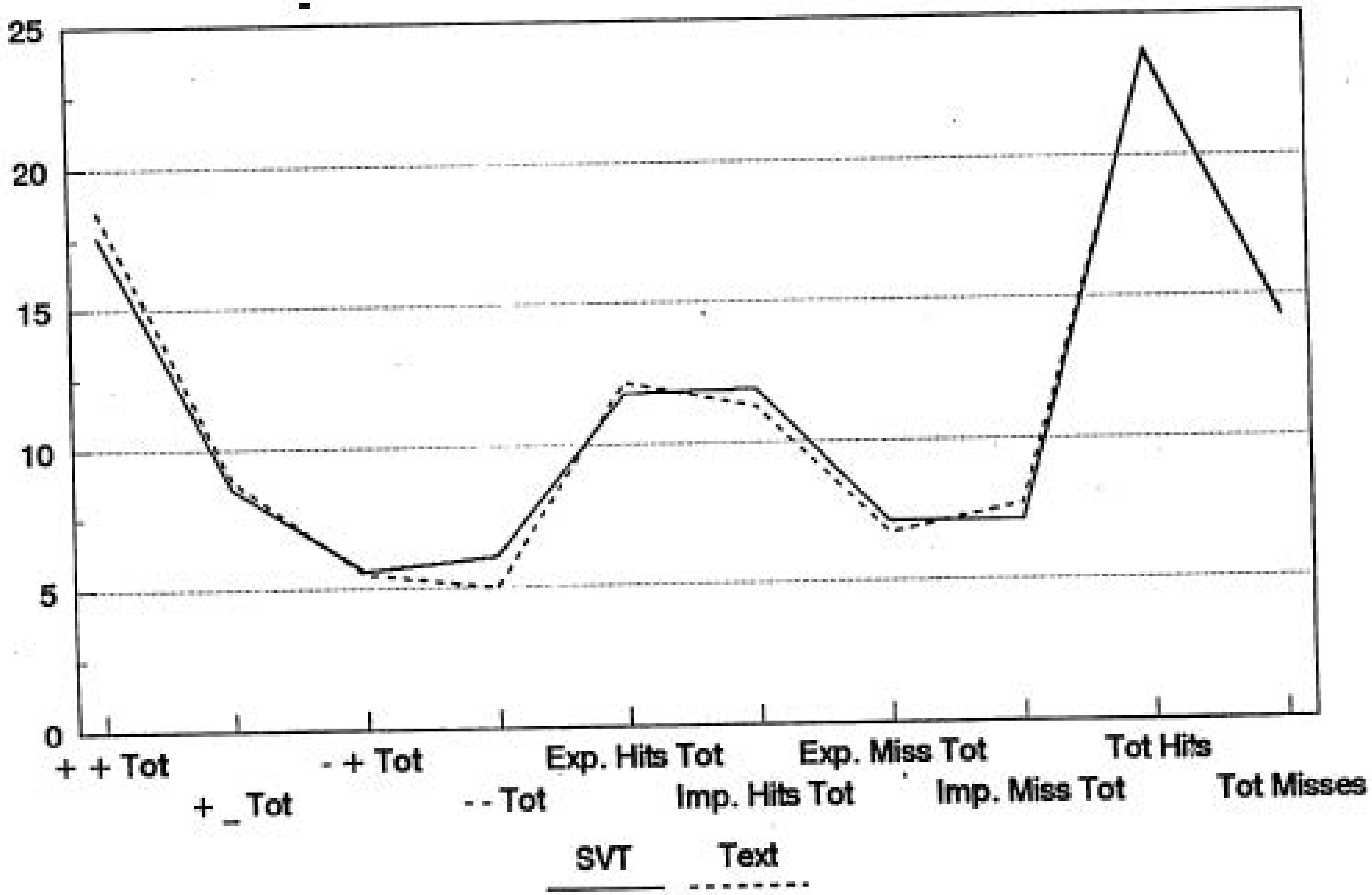


Figure 5.

High and Low Metacognition Differences

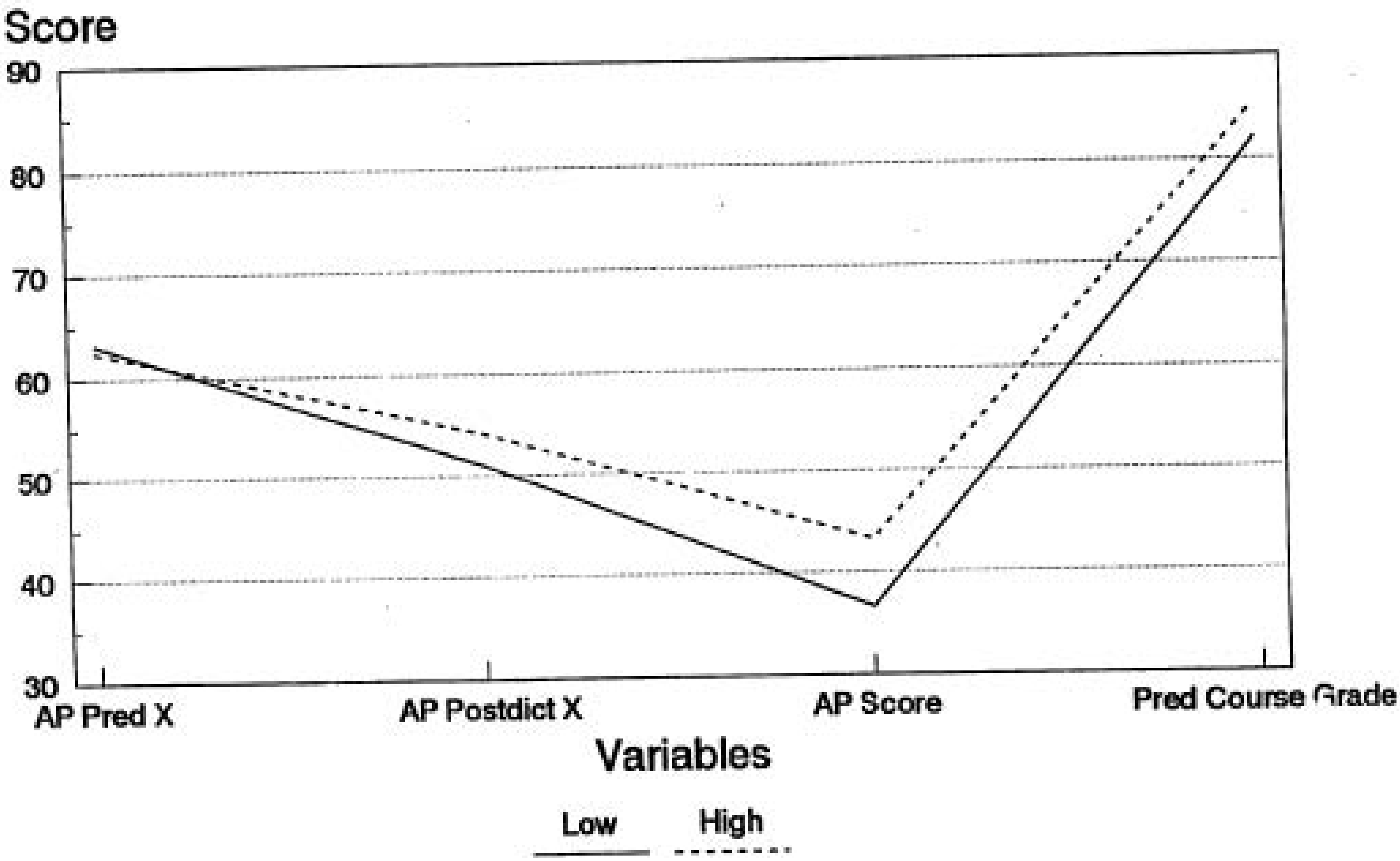


Figure 6