CollegeBoard
connect to college success™

Research Report
No. 2002-3

Knowing What You Know and What You Don't: Further Research on Metacognitive Knowledge Monitoring

Sigmund Tobias and Howard T. Everson

www.collegeboard.com

College Board Research Report No. 2002-3

Knowing What You Know and What You Don't: Further Research on Metacognitive Knowledge Monitoring

Sigmund Tobias and Howard T. Everson

College Entrance Examination Board, New York, 2002

Sigmund Tobias is a distinguished scholar in the Division of Psychological and Educational Services in the Graduate School of Education at Fordham University at Lincoln Center.

Howard Everson is Vice President for Academic Initiatives and Chief Research Scientist at the College Board.

Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board: Expanding College Opportunity

The College Board is a national nonprofit membership association dedicated to preparing, inspiring, and connecting students to college and opportunity. Founded in 1900, the association is composed of more than 4,200 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 22,000 high schools, and 3,500 colleges, through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of equity and excellence, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, contact www.collegeboard.com

Additional copies of this report (item #993815) may be obtained from College Board Publications, Box 886, New York, NY 10101-0886, 800 323-7155. The price is $15. Please include $4 for postage and handling.

Copyright © 2002 by College Entrance Examination Board. All rights reserved. College Board, Advanced Placement Program, AP, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board. PSAT/NMSQT is a registered trademark jointly owned by both the College Entrance Examination Board and the National Merit Scholarship Corporation. Visit College Board on the Web: www.collegeboard.com.

Printed in the United States of America.

Acknowledgments

This work was funded by the College Board. Earlier versions of this paper, and the studies reported herein, were presented in April 1999 at the annual meeting of the American Educational Research Association in Montreal. The findings and opinions expressed in this article are our own and do not reflect the positions or policies of the College Board.

The research reported here was conducted in collaboration with our colleagues: Lourdes Fajar and Katherine Santos conducted Study I in partial fulfillment of the requirements for the Seminar in Educational Research at the City College of New York; Rhonda Romero conducted Study II in partial fulfillment of the requirements for the Seminar in Educational Research at the City College of New York; Edgar Feng conducted Study III in partial fulfillment of the requirements for the Seminar in Educational Research at the City College of New York; Fred Dwamena conducted Study IV in partial fulfillment of the requirements for the Seminar in Educational Research at the City College of New York; Study V was supported by a contract with the Battelle Corporation relying on resources made available by the Navy Personnel Research and Development Center; Harold Ford conducted Study VII in partial fulfillment of the requirements for the qualifying examination at Fordham University's Graduate School of Education; Julie Nathan conducted Study X as her doctoral dissertation research at Fordham University's Graduate School of Education; Hyacinth Njoku conducted Study XI in partial fulfillment of the requirements for the qualifying examination at Fordham University's Graduate School of Education.

We are grateful to both Patrick C. Kyllonen and Irving Katz of Educational Testing Service for their helpful comments in earlier versions of this paper.

Contents

I. Introduction

For more than a decade our program of research has concentrated on furthering our understanding of one aspect of metacognition—knowledge monitoring. Our research has been animated by a desire to understand learners' ability to differentiate between what they know and do not know. In general, metacognition, perhaps the most intensively studied cognitive process in contemporary research in developmental and instructional psychology, is usually defined as the ability to monitor, evaluate, and make plans for one's learning (Brown, 1980; Flavell, 1979). Metacognitive processes may be divided into three components: knowledge about metacognition, monitoring one's learning processes, and the control of those processes (Pintrich, Wolters, and Baxter, 2000). We believe that monitoring of prior learning is a fundamental or prerequisite metacognitive process, as illustrated in Figure 1.

If students cannot differentiate accurately between what they know and do not know, they can hardly be expected to engage in advanced metacognitive activities such as evaluating their learning realistically, or making plans for effective control of that learning.

To date we have completed 23 studies of knowledge monitoring and its relationship to learning from instruction. Our earlier work, 12 studies in all, is summarized and reported elsewhere (see Tobias and Everson, 1996; Tobias and Everson, 2000). In this paper we continue this line of research and summarize the results of 11 studies that have been conducted over the past three years. The work reported here attempts to address a number of general issues, e.g., the domain specificity of knowledge monitoring, measurement concerns, and the relationship of knowledge monitoring to academic ability. In addition to suggesting new directions for further research, we also discuss the implications of this research for learning from instruction.

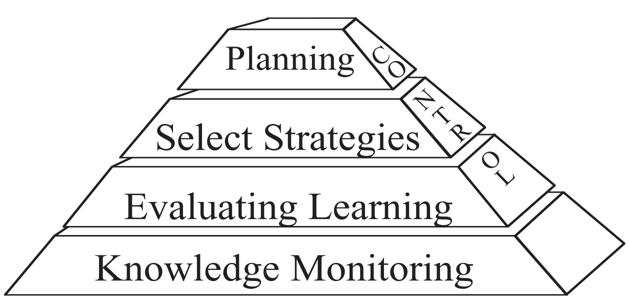

Figure 1. A componential model of metacognition.

The Importance of Knowledge Monitoring

Our interest in the accuracy of monitoring prior knowledge stems from our belief that this ability is central to learning from instruction in school and in training settings in business, industry, and the government (Tobias and Fletcher, 2000). Learners who accurately differentiate between what has been learned previously and what they have yet to learn are better able to focus attention and other cognitive resources on the material to be learned. Much of the research conducted to date supports this supposition.

Our earlier research, for example, indicated that knowledge monitoring ability was related to academic achievement in college (Everson, Smodlaka, and Tobias, 1994; Tobias, Hartman, Everson, and Gourgey, 1991). Moreover, the relationship between knowledge monitoring and academic achievement was documented in diverse student populations, including elementary school students, students attending academically oriented high schools, vocational high school students, college freshmen, and those attending college for some time. Again, details of these studies can be found in our earlier reports (Tobias and Everson, 1996; 2000). More recently, we have concentrated on the development of knowledge monitoring assessment methods that can be used across academic domains, and that have measurement properties which allow for greater generalizability of results.

II. Assessing Knowledge Monitoring

Metacognitive processes are usually evaluated by making inferences from observations of students' performance, by interviewing students, or by self-report inventories. As Schraw (2000) noted, developing effective methods to assess metacognition has been difficult and time consuming. Such assessments usually require detailed observation and recording of students' learning, rating the observations for metacognition, obtaining "think aloud" protocols of students' work, and rating their introspective reports. Referring to this approach, Royer, Cisero, and Carlo (1993) noted that "The process of collecting, scoring, and analyzing protocol data is extremely labor intensive" (p. 203). Obviously, self-report scales would be the most convenient tools to measure metacognition, and a number of such questionnaires have been developed (Jacobs and

Paris, 1987; Pintrich, Smith, Garcia, and McKeachie, 1991; Schraw and Denison, 1994; Tobias, Hartman, Everson, and Gourgey, 1991). Self-report instruments have the advantage of easy administration and scoring. However, their use raises a number of questions which have been detailed elsewhere (Tobias and Everson, 2000) and will not be summarized here.

In contrast, the knowledge monitoring assessment (KMA) technique developed for use in our research program evaluates the differences between the learners' estimates of their procedural or declarative knowledge in a particular domain and their actual knowledge as determined by performance. The accuracy of these estimates is measured against test performance. This approach is similar to methods used in research on metamemory (Koriat, 1993; Nelson and Nahrens, 1990), reading comprehension (Glenberg, Sanocki, Epstein, and Morris, 1987), and psychophysics (Green and Swets, 1966). A review of research on existing metacognitive assessment instruments (Pintrich et al., 2000) found that the scores on the KMA had the overall highest relationship with learning outcomes.

Analysis of Monitoring Accuracy

Clearly, with the KMA we are concerned with assessing knowledge monitoring ability, i.e., the accuracy of the learners' knowledge monitoring. In this measurement framework, the data conform to a 2 X 2 contingency table with knowledge estimates and test performance forming the columns and rows. In our earlier research we reported four KMA scores for each student which provided a profile of their knowledge of the domain and whether they demonstrated that knowledge on a subsequent test. The (+ +) and the (- -) scores were assumed to reflect accurate knowledge monitoring ability, and the (+ -) and (- +) scores inaccurate knowledge monitoring. A number of scholars, including Nelson (1984), Schraw (1995), and Wright (1996), have suggested that the optimal analysis of the discrepancies between estimated and demonstrated knowledge requires a probabilistic conceptualization of knowledge monitoring ability. They encourage the use of either the Gamma (*G*) coefficient, a measure of association (Goodman and Kruskal, 1954), or the Hamman coefficient (*HC*), a measure of agreement accuracy (Romesburg, 1984). These and similar methods have been used in metamemory research on the feeling of knowing and judgments of learning (Nelson, 1984).

Though there is some debate about which measure is more suitable (Wright, 1996), Schraw (1995) has argued that *G* is less appropriate when the accuracy of agreement is central, as it is in the KMA paradigm. Schraw (1995) demonstrates, and our work supports this assertion (Tobias, Everson, and Tobias, 1997) that calculating *G* may actually distort the data and lead to different inferences of ability. This can be seen, for example, in Table 1, below, which displays two hypothetical KMA score patterns where accuracy of agreement is equivalent, but the distributions across the 2 x 2 table differ (i.e., 10 accurate and five inaccurate knowledge estimates). The *G* coefficients differ, .61 and .45, while the *HC*'s are identical, .33 for each. In our earlier work (Tobias et al., 1997) we found identical *G*'s even though the knowledge monitoring accuracy differed, whereas the *HC*'s were different for these score distributions. A major disadvantage of the *G* arises when any of the 2 X 2 cells are empty, *G* automatically becomes 1. *HC* estimates, on the other hand, are unaffected by empty cells in the score distributions. Since there are often a number of empty cells in the response patterns in our research, the utility of using *G* as an estimator is questionable.

In view of these considerations, Schraw (1995) suggested using both *G* and *HC*. Wright (1996) and Nelson and Nahrens (1990) have pointed out that the *HC* is dependent on marginal values and can, therefore, lead to inaccurate assessments of the estimate–performance relationship. Such problems arise when all possible combinations of estimates and performance are considered, i.e., when all four cells of the 2 X 2 table are of equal interest. Since we are concerned only with the accuracy of estimates, or the agreement between estimates and test performance, the *HC* coefficient appears to be the most useful statistic for the analyses of these data. The *HC* coefficients range from 1.00, signifying perfect accuracy, to -1.00, indicating complete lack of accuracy; zero coefficients signify a chance relationship between estimated and demonstrated knowledge. Further support for using *HC* comes from two studies reported below (Studies I and II). In these studies the correlations between *HC* and *G* averaged .85, suggesting that using *HC* would not provide biased estimates of knowledge monitoring accuracy. Thus, *HC* was used in the studies reported below as an estimate of knowledge monitoring accuracy.

TABLE 1

Two Prototypical KMA Item Score Patterns

Test	*Knowledge Estimates*			
Performance	*Know*	*Don't Know*	*Know*	*Don't Know*
Pass	5	3	8	3
Fail	2	5	2	2

III. Knowledge Monitoring Accuracy and the Developing Reader

The substantial relationship between knowledge monitoring accuracy and reading comprehension measures of college students (Tobias et al., 1991; Everson et al., 1994) suggests that knowledge monitoring accuracy and reading comprehension ought to have similar relationships at all educational levels. Of course, good reading comprehension is especially important in elementary school since students who fall behind early in school have a difficult time catching up. Therefore, the goal of the first two studies was to examine the knowledge monitoring–reading relationship among young elementary school students.

Study I: Knowledge Monitoring Accuracy and Reading in Bilingual Elementary School Students[1]

This study examined the differences in knowledge monitoring accuracy between mono- and bilingual students, as well as the relationship between this metacognitive ability and reading comprehension in relatively young school children. Jimeniez, Garcia, and Pearson (1995) found that when bilingual students come upon an unfamiliar English word they often search for cognates in their native language. They also reported that bilingual students, when compared to their monolingual peers, monitored their comprehension more actively by asking questions when they faced difficulties or by rereading the text. This suggests that bilingual children attempting to comprehend text presented in English are likely to be more accurate knowledge monitors than their monolingual peers. This hypothesis was tested in Study I.

Participants and Procedures

Fifth- and sixth-grade students (n=90) from two large, urban public schools participated in this study. Two-thirds of the participants were bilingual, reporting that Spanish was their first language. Knowledge monitoring accuracy was assessed using the standard KMA procedure, i.e., a 34-item word list was presented, and the students indicated the words they thought they knew and those they did not. A multiple-choice vocabulary test that included the words presented in the estimation phase followed. The vocabulary words were selected for grade-level appropriateness, and were presented in order of increasing difficulty. The word list and vocabulary test were also translated into Spanish. The 60 bilingual participants were randomly assigned to one of two groups: a group tested with a Spanish language KMA, and a group tested with an English language KMA. The monolingual group, serving as a contrast group, also took the English version of the KMA. Archival measures of reading ability based on performance on the *Degrees of Reading Power* (DRP) test (Touchstone Applied Science Associates, 1991) were retrieved from school's files.

Results and Discussion

Participants were divided into high and low reading ability groups using a median split on the DRP to create the groups. A 3 x 2 ANOVA, three language groups and two reading ability groups, was conducted with the *HC* derived from the KMA as the dependent variable. Differences across the three language groups on the KMA were not significant (F [2,84]=< 1). Good and poor readers, however, did differ (F [1,84]=6.56, $p < .01$), with the better readers demonstrating higher metacognitive monitoring ability. The interaction between language groups and reading ability was not significant. The finding that good readers were more accurate monitors than the poorer readers fits with earlier research (Tobias and Everson, 2000). However, the magnitude of the knowledge monitoring–reading relationship was somewhat lower for the school age students (r=.28) than for college students (r=.67).

The absence of knowledge monitoring differences between mono- and bilingual students may be attributed to the English language fluency of the bilingual group. Subsequent interviews with the bilingual students indicated that the majority lived in the United States for four or more years and were fluent in English. Such fluency, apparently, made it unnecessary to monitor comprehension and search for cognates in their native language.

Study II: Reading, Help Seeking, and Knowledge Monitoring[2]

Research evidence points to the fact that good readers may be more aware metacognitively than poorer readers. Though our prior research has been concerned with knowledge monitoring accuracy and its

[1] This report is based on a paper by Fajar, Santos, and Tobias (1996) found in the references.

[2] This report is based on a paper by Romero and Tobias (1996) found in the references.

relationship to learning outcomes, we have not examined the influence of accurate knowledge monitoring on the processes invoked while learning from instruction, i.e., the components of metacognition that control learning from instruction.

The intent of Study II was to examine one such process, help seeking—an important learning strategy when one is baffled, confused, or uncertain while trying to learn something new or when solving novel problems. Seeking help, we argue, signals a level of metacognitive awareness—a perceived gap in knowledge, perhaps—and an intent on the part of the learner to address the learning problem. Achieving that awareness suggests that learners can differentiate between what they know and do not know. Thus, we hypothesized that measures of knowledge monitoring ability should correlate with help seeking activities in the reading comprehension domain. Simply put, accurate knowledge monitors should seek help strategically, i.e., on material they do not know because soliciting help on known content wastes time that could be spent more usefully seeking assistance on unknown content. Less accurate monitors, on the other hand, are unlikely to be strategic and were expected to seek more help on known materials.

Participants and Procedures

Forty-one fourth-grade students (49 percent male) from an urban public school participated. They were ethnically diverse, and a number of the students reported they were from families with incomes below the poverty line. The participants were, for the most part, selected from regular elementary school classes, though four (10 percent) were mainstreamed into the classes from special education.

As in our earlier studies, the KMA consisted of a 38-item word list and vocabulary test generated from fourth-grade curriculum materials. Participants' scores on the DRP (Touchstone Applied Science Associates, 1991) were obtained from school records. Help seeking was operationalized by asking participants to leaf through a deck of 3 X 5 index cards containing the 38 words appearing on the KMA and select 19 for which they would like to receive additional information. The information, printed on the back of each index card, consisted of a definition of the word and a sentence using the word in context. Participants were tested individually and the words selected for additional help were recorded.

Results and Discussion

As expected, the correlation between the KMA and DRP scores was .62 ($p < .001$), which was quite similar to the correlation of ($r = .67$) found for college students (Tobias et al., 1991). While the DRP scores for this sample were somewhat more variable than in Study I, it is not clear why the correlations were substantially higher than in the earlier study. It is plausible that the archival DRP scores of the bilingual students in Study I were not representative of their developing reading abilities. Despite this variation in the correlations, the results of these two studies indicate that the metacognitive monitoring–reading relationship is similar at both the elementary and postsecondary levels.

To analyze help seeking behavior we split the participants into high and low knowledge monitoring groups, and four-word category groups: (1) words known and passed on the test (+ +); (2) words claimed as known but not passed (+ -); (3) words claimed as unknown and passed (- +); and (4) words claimed as unknown and not passed (- -). The dependent measures were derived by calculating the percent of words selected by the participants for further study from each of the four-word categories. A 2 x 4 ANOVA with repeated measures on the second factor was computed. As expected, a highly significant difference among the four-word categories was found (Wilks F [3,37]=36.22, $p < .001$). More important, a significant interaction was found between knowledge monitoring accuracy and word types studied (Wilks F [3,37]=15.34, $p < .001$). This interaction is displayed in Figure 2.

The results indicate that participants with higher KMA scores asked for more help, by a small margin, on words estimated to be unknown and failed on the test (- -), whereas those with lower KMA scores asked for help more often with words estimated to be known and failed on the test (+ -). There was one exception to that trend: those with higher KMA scores also sought more help on the words they claimed they knew and did, in fact, know (+ +). This was clearly not a strategic use of help, and a waste of their time. Upon reflection, it is

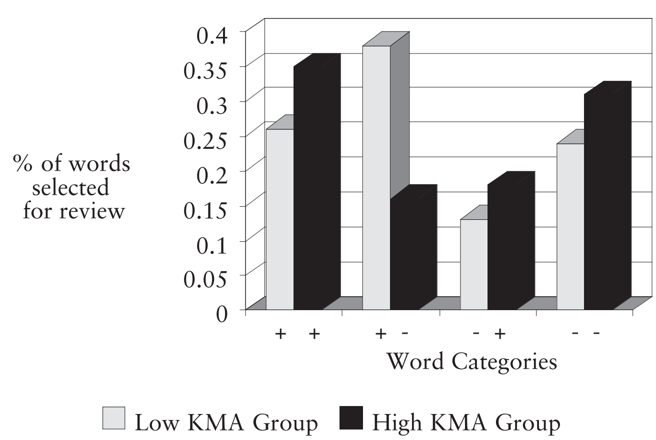

Figure 2. The interaction between knowledge monitoring ability and help seeking reviews.

plausible that one reason for seeking help on known items was that students were asked to review a fixed number of words. Some participants, for example, commented that they would have reviewed fewer words had they been allowed. Thus, by requiring participants to review a fixed number (19) of words, we may have confounded the findings. In Study III, below, we attempted to modify this design flaw.

Summary: Knowledge Monitoring and Reading

The results of the two studies of the reading–knowledge monitoring relationship among elementary school students indicate that this aspect of metacognition is important for school learning. The findings of both studies indicated that there were significant positive relationships between metacognitive knowledge monitoring and reading. These findings comport with the results of our earlier work (Tobias and Everson, 1996; 2000), even though there was some variability in the magnitude of the reading–knowledge monitoring relationship between the two studies reported here. The variability may be attributable to the small samples used in Studies I and II, or to the possibility that the reading and metacognition relationships are more variable for elementary school students (i.e., developing readers) than for more mature college students. Further research with larger, developmentally varied samples is needed to clarify this issue.

Study III: Strategic Help Seeking in Mathematics[3]

This investigation extended Study II by examining help seeking behavior in mathematics. We expected that relatively accurate knowledge monitors would seek more help on math problems they thought they could not solve than their less accurate peers who were expected to be less strategic and less metacognitively able.

Participants and Procedures

A total of 64 tenth-grade minority students (33 females) participated in the study. They were administrated 26 mathematics problems, divided evenly between computation and problem solving, selected from prior New York State Regents Competency Tests. Participants had to estimate whether they could, or could not, solve each problem and then received a test asking them to solve the same problems. After the test, participants selected problems on which they wished to receive additional help. The help consisted of the correct answer and the steps required to solve the problems. The help was printed next to the problem and covered so that it could not be seen without removing stickers covering the help material. Participants were asked to remove the stickers when they wished additional help on the problem.

Results and Discussion

Participants were divided into high and low knowledge monitoring groups. The mean percentages of help sought for each group on four types of problems were determined. The item types were (1) those estimated as solvable and solved correctly (+ +); or (2) those estimated as solvable and solved incorrectly (+ -); or (3) those estimated as unsolvable and solved correctly (- +), or (4) those estimated as unsolvable and failed on the test (- -). These data are shown in Figure 3.

The data were analyzed using multivariate analysis of variance (MANOVA). There were no differences in the mean number of problems on which both groups chose additional help, but differences emerged on the type of problems for which help was sought (Wilks F [4,59]=3.71, $p < .01$). As Figure 3 indicates, the largest differences between accurate and less accurate knowledge monitors occurred on the help sought for those problems participants originally estimated as solvable and then failed on the test. (F [1,62]=10.36, $p < .01$). Seeking help on these items is of strategic value, since participants are likely to have realized they were wrong to estimate they could solve these problems since they

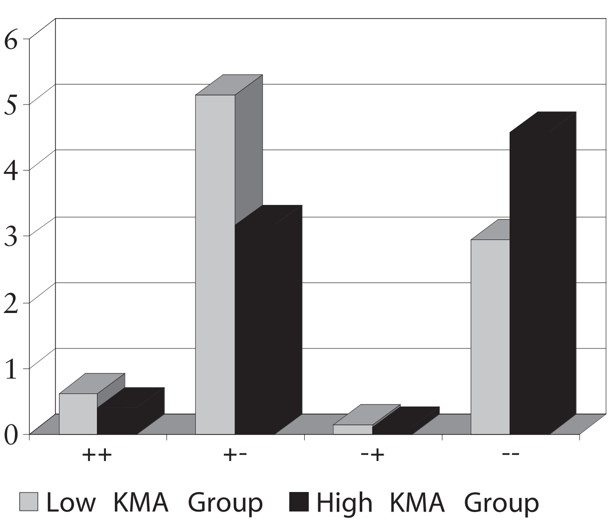

Figure 3. The relationship between knowledge monitoring and help seeking reviews in Study III.

[3] This report is based on a paper by Tobias and Everson (1998) found in the references.

subsequently failed them on the test. Therefore, it made good sense to seek help on these items in order to determine the reasons for their wrong estimates. Another difference was on problems perceived as unsolvable that were failed subsequently on the test (F [1,62]=8.62, p=< .03). This finding is similar to Study II and indicates that the metacognitively accurate students reviewed more strategically than their less accurate peers. None of the other differences were significant.

In this study help was sought on far fewer problems (36 percent overall) than on Study II where students were required to review half the problems. Therefore, it seems clear that students were required to seek more help in the earlier study than was needed which may have affected those results. Thus, the differences between the two studies may be greater than the percentages suggest since there are two reasons for expecting students to seek more help on mathematics than on vocabulary items. First, solving mathematics problems is usually considered to be a more difficult task than finding vocabulary synonyms. Second, the assistance offered in Study III was greater than for the earlier study since students had more control over the number of terms to review, and then could see the steps required to solve the problem.

There were also differences in the types of items on which help was sought in the two studies. A comparison of Figures 2 and 3 indicates that in Study III far less help was sought on problems students estimated knowing and passed on the test, as well as those estimated as being unknown and passed on the test. Apparently, not requiring participants to seek help on a specific number of items led to a reduction in the number of problems reviewed and also allowed participants to be more selective in their help seeking.

It is surprising that no correlations were found between KMA scores and grades on in-class mathematics examinations. One reason may be that the participants were permitted to use calculators during in-class mathematics tests but not on the study materials; many complained about being denied the use of calculators. Nevertheless, the findings of Study III extend prior research on the relationship of knowledge monitoring to students' strategic choices in seeking additional assistance for learning. Further research with larger samples is clearly needed to examine the different types of strategic behavior of the two knowledge monitoring groups. Study III also confirms the relationship between knowledge monitoring in mathematics and learning in that domain. Strategic help seeking among secondary school students was examined in Studies X and XI and will be discussed later in this report.

IV. Knowledge Monitoring and Ability

Metacognition is described by a number of researchers as an executive process (Borkowski, 2000) that directs the functioning of other cognitive processes. In Sternberg's (1988) theory of intelligence a critical role is played by general cognitive processes, or metacomponents that are analogous to metacognition, which coordinate the functions of lower order cognitive processes used in planning, monitoring, and evaluating learning. These formulations clearly suggest that metacognition, in general, and knowledge monitoring, in particular, should be related to measures of intelligence and learning ability. Prior research (Tobias and Everson, 2000), and the studies described above, indicate that accurate monitoring is related to reading, achievement in mathematics, and learning as reflected in school grades. These findings suggest that knowledge monitoring should also be related to different measures of learning ability or intelligence, though none of our earlier studies dealt directly with that question. One purpose of the four studies described below was to investigate more directly the knowledge monitoring–ability relationship.

Study IV: Triarchic Intelligence and Knowledge Monitoring[4]

Sternberg (1995) has maintained that tests of intelligence or scholastic aptitude measure analytic abilities related to success in most levels of schooling. However, he maintains that analytic ability tells us little about the types of processes needed to solve problems in everyday life, such as practical intelligence or creative abilities. Sternberg's triarchic theory of intelligence (1995) contends that measures of both practical and creative abilities should be incorporated into intelligence tests in order to predict success in a range of contexts beyond school. Sternberg suggests that intelligence is demonstrated in figural content areas and not only in the verbal and mathematical arenas frequently assessed in traditional measures of intelligence, and he has designed tests to assess each of these domains. In view of the prior relationships between knowledge monitoring and school learning it was expected that knowledge monitoring should be related to verbal analytic intelligence, especially when the KMA is composed of vocabulary tasks. Further,

[4] We are grateful to Robert Sternberg for making these tests available to us.

we were interested in the KMA's relationships to measures of practical and creative intelligence.

Participants and Procedures

Sixty participants were recruited from a high school in a low socioeconomic community attended largely by minority students. A word list was developed from tenth-grade curriculum materials, and participants were asked to check whether they knew or did not know each of the words and then complete a multiple-choice vocabulary test based on the same words. Participants also received these nine tests from Sternberg's (1995) triarchic intelligence battery: (1) analytic ability, (2) practical intelligence, and (3) creative intelligence. Three tests using verbal, quantitative, and figural stimuli for each type of intelligence posited by Sternberg were administered, yielding a total of nine tests that could be group administered and objectively scored.

Results and Discussion

Knowledge monitoring accuracy was correlated with students' scores on the nine subtests from Sternberg's battery. Of the nine correlation coefficients only two were significant: the analytic verbal subtest had a correlation of .40 ($p < .01$) with the KMA, while the practical quantitative test correlated .29 ($p < .05$). The significant correlation with analytic verbal ability was expected since prior research (Tobias and Everson, 1998; 2000) had shown that vocabulary based KMAs were related to school learning, the primary indicator of analytic ability. However, it is difficult to account for the relationship with practical quantitative ability, especially since the KMA was based on a vocabulary test rather than on mathematics problems. Since eight correlations were computed for this analysis on an exploratory basis, the possibility that one of these was attributable to chance cannot be ignored, and replication is needed in order to establish further the relationship between knowledge monitoring and diverse measures of intelligence.

Study V: The Impact of Knowledge Monitoring, Word Difficulty, Dynamic Assessment, and Ability on Training Outcomes[5]

Our prior research used students from educational institutions, elementary through college, as participants. Study V investigated the knowledge monitoring–ability relationship of adults in a training context. The ability measures used in this study were developed by the U.S. armed forces and were not designed for school settings. Two other issues were also addressed in this study: the impact of word difficulty and dynamic assessment procedures on knowledge monitoring. It was expected (Tobias and Everson, 2000) that medium difficulty vocabulary words would provide the most accurate index of knowledge monitoring. Responses to easy words were assumed to be relatively automatic, requiring minimal reflective processing by participants; difficult words, on the other hand, were likely to be dismissed as too complex and most likely to receive little thoughtful attention. Words of medium difficulty, in contrast, were expected to invoke the most intense processing since a majority of trainees could perceive themselves as identifying the correct definition of these words. Therefore, moderately difficult words were expected to produce the most useful indices of the discrepancy between estimated and demonstrated knowledge.

In previous studies (Tobias and Everson, 2000), a text passage was used in which all of the words on the KMA were defined. In this study we expected that enabling trainees to infer word meaning from the text increased the similarity to dynamic assessment approaches (see Carlson and Wiedl, 1992; Guthke, 1992; Lidz, 1992) in which learners have the opportunity to learn before and during testing. Previous results indicated that relationships with reading comprehension were substantially higher for those given the chance to infer the meanings of words (Tobias et al., 1991). However, reading a text covering the KMA words had little impact on the relationship between knowledge monitoring and college grades (Everson and Tobias, 1998). Study V examined whether a text passage, in which some of the words used in the KMA were defined, might be more closely related to training outcomes where the objectives of instruction were more explicit than in most educational contexts (Tobias and Fletcher, 2000). Higher relationships with learning outcomes were expected when the monitoring assessment and the instruction shared a similar content and context.

Participants and Procedures

The 71 participants (20 percent female) were examined in groups of 12–18 at a U.S. Navy training school in San Diego, California. The dependent variables consisted of participants' mean scores, provided by the school, on six multiple-choice tests taken during the six-week course. During the course participants were trained to apply concepts from the physics of

[5] This study was supported by a grant from the Navy Personnel Research and Development Center. This summary is based on a paper by S. Tobias, H.T. Everson, and L. Tobias (1997).

underwater sound transmission to the task of locating and tracking submarines underwater using sophisticated sonar equipment.

Materials

All participants received an 80-item word list and had to estimate whether they knew or did not know each word. The list contained 40 general words, obtained from prior versions of the Vocabulary subtest of the Gates-MacGinitie Achievement[6] battery for secondary school students. Item difficulty data from this test indicated there were 13 easy items, 14 medium items, and 13 difficult items. Another 40 words dealt with the field of oceanography and were composed of two subgroups: 15 technical words (e.g., *benthic, guyot, nektons*) and 25 more common words which were generally related to the domain in which students were receiving training (e.g., *flora*, *habitats*, *marine ecology*). A multiple-choice vocabulary test containing the same words was then given before students read a three-page text passage on oceanography, methods of exploring the ocean, and the characteristics of underwater environments. All the oceanography words were defined either explicitly or implicitly in the text. The word list and vocabulary test dealing with oceanography were then readministered in order to obtain an index of students' ability to update their subject specific knowledge and monitoring ability.

Results and Discussion

The results dealing with KMA–ability relationships, and the impact of knowledge updating will be discussed here; data dealing with the domain generality/specificity issue will be discussed later in the report. Scores on the Armed Forces Qualification Test (ASVAB) and on selected subtests of the ASVAB were available for a significant number of students. These scores were correlated with the KMA for each type of word; the results and number of cases for each correlation are shown in Table 2. The results indicate that, as expected, there were significant correlations between the KMA and those subtests of the ASVAB measuring verbal ability.

Contrary to expectations, the highest knowledge monitoring–ability relationships were found for easy words; correlations decreased for more difficult and more technical words. These findings suggest that the influence of general ability on knowledge monitoring decreased as more difficult and more specialized knowledge was needed to make accurate knowledge monitoring estimates. It is also important to note that correlations between the KMA and ability measures were higher after students read the oceanography text passage. Apparently, the ability to learn new words by inferring their meaning from the text was more heavily associated with general ability than was prior familiarity with the technical words.

The means and standard deviations of the KMA scores for all types of words are shown in Table 3, as are the descriptive data for the mean scores on all tests given in the course. A within-subjects ANOVA indicated that the differences among the KMA scores were significant ($F\ [6,420]=39.66$, $p < .001$).

Again, monitoring accuracy declined as the difficulty and technical content of the words increased. The results also indicated that KMA scores increased on the second administration after the text was read, presumably because the trainees recalled, or were able to infer, the definition of the KMA words from the passage.

The impact of monitoring accuracy on learning was examined further by splitting the ASVAB test scores at the median and computing a 2 (low/high test score) X 7 (word types) ANOVA with repeated measures on the second factor; the KMA scores served as the dependent variable. As expected, higher ability students were more accurate knowledge monitors ($F\ [1,69]=5.28$, $p < .05$), see Figure 4. The interaction of word type and ability test score was not significant.

In sum, results suggest that knowledge monitoring was related to the trainees' ability to learn from the

TABLE 2

Correlations Among ASVAB Measures and KMA Scores by Word Difficulty

	KMA Scores by Word Categories						
ASVAB Measures	*Easy*	*Medium*	*Difficult*	*Common*	*Technical*	*Common 2*	*Technical 2*
AFQT (67)	.59**	.40**	.33*	.44**	.20	.55**	.35**
Word Know (43)	.68**	.40**	.23	.36*	-.02	.52**	.26
Para. Comp. (38)	.59**	.34*	.23	.12	.22	.28	.28
Verbal (33)	.77**	.48**	.33	.33	.15	.59*	.43*
Gen. Science (49)	.44**	.29*	.33*	.42*	.17	.25	.40*

*p < .05
**p < .01

[6] We are grateful to the Riverside Publishing Company for providing these data for our research.

TABLE 3

Descriptive Statements for KMA Scores and GPA

Variables	*Mean*	*SD*	*Minimum*	*Maximum*
Easy	.51	.12	-.10	.62
Medium	.36	.14	.00	.55
Difficult	.28	.18	-.24	.62
Common	.21	.17	-.21	.58
Technical	.21	.15	-.30	.55
Common 2	.30	.18	-.09	.64
Technical 2	.37	.17	-.04	.65
GPA (M of 6 tests)	89.23	4.59	78	99

N=71 for all cells

passage and with their learning in the Navy training course. These results were similar to earlier findings (Everson and Tobias, 1998; Tobias and Everson, 1996; 2000) indicating that the KMA was related to learning among college, high school, and vocational high school students. Since the interaction between word types and outcomes was not significant, presumably because of the relatively small number of cases, selected comparisons on specific word types were not computed. We expected that words dealing with oceanography would be most highly related to instructional outcomes. The pattern of results tends to confirm these expectations, though they did not reach statistical significance. The largest differences between high and low achievers occurred on the common oceanography words before and after reading the passage, and on the technical words once the text was read, apparently reflecting the importance of accurate knowledge monitoring in learning this new material. We also expected that the largest monitoring differences for words not dealing with oceanography would be for items of medium difficulty. The pattern of results suggests that differences on the easy, medium, and difficult words were relatively similar, failing to confirm our earlier predictions about word difficulty.

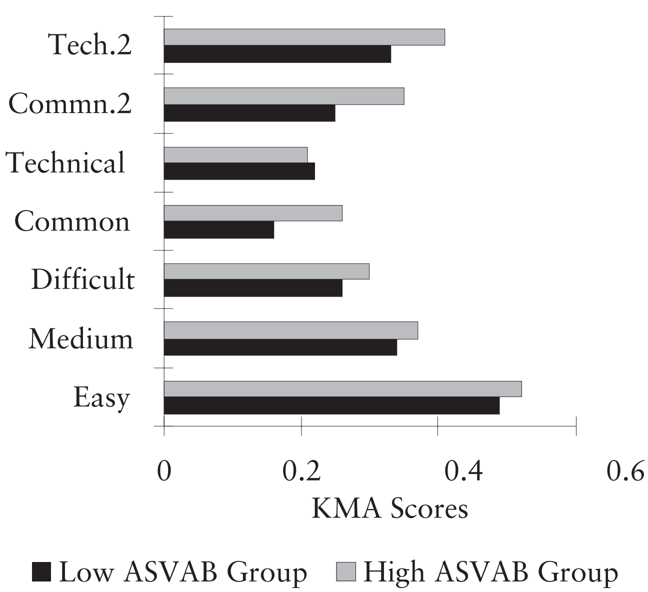

Figure 4. Relationship of KMA scores with ASVAB performance.

The data in Table 3 indicate that the trainees' test scores were relatively homogenous, and perhaps attributable to the mastery based organization of the course (i.e., students who failed a test by obtaining scores less than 70 percent had to restudy the materials and take another test, restricting the achievement range and potential relationships of training outcomes to external criteria). In mastery based courses the time devoted to learning may be related more to external measures, such as the KMA, than are grades, and one recommendation for further research is that learning time data be collected in studies using mastery learning. Nonetheless, these results suggest that the highest relationships between the KMA and learning outcomes occurred on all oceanography words after students had the opportunity to update their word knowledge. This is contrary to prior research (Tobias and Everson, 2000) that found the highest relationships between the KMA and grade-point averages in college *before* students had the chance to update their word knowledge. Presumably, the similarity of the oceanography passage to the instructional domain, underwater acoustics, may account for the facilitating effect, and it may be that metacognitive knowledge monitoring is most useful in instructional situations where the content is less familiar to the learner.

Giving students the opportunity to update their word knowledge by reading the text passage approximated dynamic assessment procedures (see Carlson and Wiedl, 1992; Guthke, 1992; Lidz, 1992). Dynamic assessment usually also includes supportive intervention in students' learning, observations of their reactions to the interventions, and evaluations of students' responses to the assistance. These additional components were not present in this study; instead students had the opportunity only to update their knowledge in a domain similar to the one which was the subject of instruction. It remains for further research to clarify whether a more active, dynamic intervention might demand more accurate, ongoing knowledge monitoring.

Study VI: Knowledge Monitoring and Scholastic Aptitude[7]

Much of our knowledge monitoring research has been conducted on relatively typical school populations. Therefore, little is known about the role of knowledge monitoring in academically talented groups. In Study VI we examined a large group of students with academic talent either in mathematics or language arts and social studies. Traditionally, academic talent has been defined almost exclusively on the basis of prior school achievement and/or scholastic aptitude. This study addressed the question of establishing a broader profile of gifted and talented students by considering the range of their metacognitive abilities, generally, and their ability to accurately assess what they know and do not know, in addition to examining the relationship of metacognitive ability to scholastic aptitude.

Participants and Procedures

The participants were 462 students from grades 7 through 11 attending a voluntary summer program for academically talented students. In a two-hour session students received a list of 30 verbal analogies selected from prior versions of the SAT® I: Reasoning Test, and were asked to estimate whether they could or could not solve each analogy question. The analogy items were projected onto a screen for 8 seconds each. Then 25 math items were projected in the same way for 20 seconds each, and students again were instructed to estimate whether they could/could not solve the problems. The math items were taken from prior versions of the SAT I: Reasoning Test and the SAT II: Subject Tests. After the estimates were completed, the analogy and math items were presented as multiple-choice tests, and students were given as much time as needed to answer the questions. Participants' recorded SAT I verbal and math scores were obtained from program archives.

Results and Discussion

As expected, SAT mathematical scores correlated .50 (p < .001) with math KMA scores, and the SAT verbal scores correlated .29 (p < .001) with the verbal KMA scores confirming expected relationships between knowledge monitoring accuracy and aptitude. These results are not surprising since the verbal and math tasks used in the KMA were adapted from prior versions of the SAT I and SAT II. The correlations found in this study are similar to those in Studies IV and V.

Study VII: Knowledge Monitoring, Scholastic Aptitude, College Grades[8]

The preceding studies found significant positive correlations between knowledge monitoring ability and scholastic aptitude. Earlier research (Everson and Tobias, 1998, Tobias and Everson, 2000) found significant relationships between college and high school grades and knowledge monitoring ability. In yet another study (Everson and Tobias, 1998), we found that KMA scores also predicted first year college grades. To bring these lines of research together, Study VII was designed to investigate whether KMA scores add incremental validity to the SAT I by increasing the prediction of college grades.

Participants and Procedures

Participants (n = 120) were first-year students attending a private college in a large urban area. The participants were enrolled in an academic skills development program designed to prepare them for college. Participants were administrated two KMAs: (1) A vocabulary KMA consisting of a 40-item word list and vocabulary test and (2) a 21-item mathematics KMA consisting of problems taken from prior versions of the PSAT/NMSQT® and SAT I. Participants were given 10 seconds to estimate whether they could or could not solve each problem; a multiple-choice math test on the same problems was then administered. Participants' grade-point averages (GPA) and their SAT I scores were obtained from their college records.

Results and Discussion

In contrast to prior findings, there were no significant relationships between students' monitoring accuracy and either their GPA or their SAT I scores. Possibly students in this sample, who were attending an academic skills preparation program, may have been graded unreliably with an emphasis on effort and less concern with achievement accounting for this unique and rather surprising finding. It appears that the KMAs used in this study, which had been previously employed with college freshmen, were too difficult for students lacking the academic skills expected of entering freshmen. This interpretation is supported by the KMA correlations with the SAT I verbal scores of .28 (p < .01) for the easy words and .38 (p < .01) for words of medium difficulty, but fell to a nonsignificant .09 for the difficult words. Apparently students made random guesses on the difficult words, reducing any systematic variance for the overall KMA. The findings that none of the correlations between either the SAT I and GPA, or between the

[7] Description of this study is based on a report by Tobias, Everson, and Laitusis (1999).

[8] This study was conducted by Harold Ford, Graduate School of Education, Fordham University.

KMA and GPA, were significant indicated that there was something very unusual about this group of participants. Therefore, it is clear that the question of whether the KMA added unique variance to the prediction of college grades should be examined with more heterogeneous student groups, rather than those composed solely of freshmen enrolled in college preparatory programs.

Summary: Knowledge Monitoring and Academic Ability

Collectively, these studies, with the exception of Study VIII, provide evidence of an important relationship between intellectual ability and knowledge monitoring accuracy. As was the case for the reading–knowledge monitoring relationship, there was some variability among the studies in the strength of the relationships. The positive ability–knowledge monitoring relationship is quite reasonable since both monitoring and scholastic aptitudes have been found to be associated with learning (Tobias and Everson, 2000) at secondary and postsecondary school levels. A remaining question is whether knowledge monitoring contributes unique variance to the prediction of school learning, above the variance accounted for by scholastic aptitude. Our attempt at answering that question was thwarted by the sampling problem in Study VIII. It remains for further research to provide satisfactory answers to that question.

V. Knowledge Monitoring: Is It a Domain General or Specific Ability?

A major issue in cognitive psychology has been whether cognitive processes are domain general or specific, i.e., if results dealing with cognitive processes apply largely to the content area in which they were studied or whether they generalize to other content domains. This question is of some importance for processes such as metacognition since it raises the issue of whether students' metacognitive processes in one content area are likely to be similar in other fields, and whether training metacognition in one domain is likely to lead to improvements in others.

Two studies (Everson and Tobias, 1998) found significant relationships between vocabulary based KMAs and GPAs in humanities and English courses but not in science or social science classes. It was reasoned that in those studies the KMA contained a sample of general words used in discourse within the larger culture. Findings of significant relationships with English and humanities grades probably reflected the fact that general familiarity with these disciplines did not require a technical vocabulary, whereas success in the sciences and social sciences does require mastery of a technical vocabulary that is unfamiliar to the average person. Therefore, the absence of KMA relationships with science and social science grades may have been attributable to the absence of technical words from those domains in the KMA. In turn, these findings suggest that metacognitive knowledge monitoring processes may be domain specific. Schraw, Dunkle, Bendixen, and Roedel (1995) reported that students' confidence judgments correlated across domains, irrespective of their performance providing qualified support for domain generality of monitoring skills. The studies reviewed below re-examined the domain generality/specificity question.

Study VIII: Math and Vocabulary KMAs, SAT® Tests, and Grades: Relationships[9]

One way of investigating the domain generality–specificity issue was to study the relationship between knowledge monitoring in a specific domain and achievement in that area. Prior research (Everson and Tobias, 1998) revealed low but significant relationships between GPA and a vocabulary based version of the KMA. However, the correlation between a mathematics version of the KMA and math grades had not been previously investigated. We examined this issue in Study VIII.

Participants and Procedures

A total of 120 undergraduates (73 females) from an urban college participated in the study. Participants were recruited from an introductory psychology course, and were awarded extra course credit for taking part in the study.

The experimental materials were administered in large group sessions. Participants received two KMAs, composed of (1) vocabulary words adapted from a prior Gates-MacGinitie Vocabulary test (Gates and MacGinitie, 1989) previously used in Study V, and (2) 21 math problems previously used in Study VII. Each of the items was presented singly using a computer projection device. Students estimated whether they knew the words or could solve the math problems and then received paper-and-

[9] This description is based on a study by Tobias, Everson, and Laitusis (1999) found in the references.

TABLE 4

Correlations of Math and Verbal KMA Score and GPA

GPA	*Math Hamann*	*Vocabulary Hamann*
English	.24*	.23*
Humanities	.26	.38**
Math	.24	.22
Science	.13	.28*
Social Science	.33**	.39**

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

pencil five-choice math and vocabulary tests containing the same items.

Results and Discussion

The correlation between the mathematics and vocabulary KMAs was .49 ($p < .001$). The generally significant correlations (see Table 4) between the KMA scores and GPAs replicated prior results (Everson and Tobias, 1998; Tobias and Everson, 2000) and indicated that the insignificant relationships found in the preceding study were attributable largely to some singular aspects of that sample.

Surprisingly the correlations of both the mathematics and vocabulary KMAs with GPA in different areas were remarkably similar (see Table 4). With the exception of grades in English, where both correlations were virtually identical, the vocabulary KMA had somewhat higher correlations with all the GPAs than the KMA based on math, though the differences were not significant. The GPAs in mathematics courses were somewhat less reliable because many participants took few or no mathematics classes, reducing the number of cases on which those relationships were based. The results of this study indicate somewhat higher relationships between knowledge monitoring accuracy across different domains than found in preceding research. Furthermore, similarity in the magnitude of the correlations of both KMAs with the different GPAs also indicates somewhat greater domain generality than do prior studies.

Study IX: Knowledge Monitoring, Reading Ability, and Prior Knowledge[10]

Prior knowledge of the content is one of the most powerful predictors of future learning (Dochy and Alexander, 1995). None of our earlier studies investigated the impact of prior learning on knowledge monitoring. We expected that students with substantial prior knowledge should be more accurate in their metacognitive knowledge monitoring in that domain than in others. One of the purposes of Study IX was to examine that expectation. In addition, the effect of reading ability was also re-examined in the context of prior knowledge.

Participants and Procedures

The participants (n=37) were college students mostly young women recruited from a large state university. Students were predominately upper juniors and seniors, and participation in the study was a part of their course requirements. Two KMAs were administered by computer: (1) vocabulary KMA consisted of 39 words previously used in Studies V, VII, and VIII, (2) KMA composed of 40 vocabulary words used in texts on learning theory (e.g., *schema*, *encode*, *extinction*, etc). Participants were initially shown a series of individual vocabulary items for eight seconds each, and asked to estimate whether they either knew or did not know each word. A computer program controlled the presentation of the KMA items and all responses were recorded by computer. After completing their estimates, participants received the same words in a multiple-choice vocabulary test without a time limit. Participants were then asked to complete a brief 11-item pretest based on a text selection dealing with learning theory which they read as a course requirement. The instructional materials appeared in a hypertext format, and participants clicked on various words that opened links for further elaboration. After the participants finished reading the hypertext passage, they took a posttest which included the 11 items on the pretest plus some others, in order to assess their understanding of the text passage. Students' performance on the test contributed to their grade in the course. The Nelson Denny reading comprehension test was also administered.

Results and Discussion

Two KMA scores were computed for each participant: one estimating knowledge monitoring accuracy for the learning theory words, and the second for the general vocabulary items. Multiple regression analysis was used to regress the learning theory posttest scores on the two KMA scores and the reading comprehension measures. Only the KMA score for the general vocabulary words contributed to the prediction of the posttest score ($r^2=.52$, $p < .001$). Table 5 shows the means and standard deviations for all the variables, the correlations of each with posttest score, and their regression coefficients.

[10] The description of this study is based on a report prepared by Tobias, Everson, Laitusis, and Fields (1999) found in the references.

TABLE 5

Descriptive Statistics, Correlations, and Beta Weights with Posttest Score

	Mean	*SD*	*r Posttest*	*Beta*
General KMA	.69	.27	.71**	.54*
Learning KMA	.17	.18	.34*	.04
Pretest scores	4.81	2.23	.34*	.08
Reading Comprehension test score	30.84	5.42	.63**	.16
Posttest score	32.38	6.43		

Pre- and Posttest $r = .34$*

* $p < .05$

** $p < .01$

It appears that the learning theory words were not as closely related to either the pretest ($r = .28$, ns) or the posttest ($r = .34$, $p < .05$) as were the general vocabulary words which were significantly related to both pretest ($r = .39$, $p < .02$) and posttest ($r = .71$, $p < .001$). Keep in mind, both pre- and posttests were referenced directly to the instructional text. On the other hand, care had been taken not to duplicate the learning theory words with terms that might appear in the instructional text in order to avoid contamination with pre- and posttest measures. That absence of duplication may have led to general KMA being more accurate predictor of learning from the instructional passage than the more domain specific.

This finding is similar to what was reported in Study V in which a KMA using technical vocabulary that was similar, but not identical to, the instructional domain differentiated between high and low achieving trainees after they had the opportunity to familiarize themselves with the vocabulary used in the KMA. Apparently, merely having a logical or conceptual connection to the subject is not sufficient to increase KMA relationships with learning in a particular domain, unless the KMA actually samples the content of the instructed domain. The results are also interesting with respect to the relative importance of prior knowledge and monitoring accuracy. Since only the general KMA scores contributed significantly to the prediction of the posttest, it suggests that monitoring accuracy may be more important for the prediction of learning than even prior knowledge of the content to be learned. The generality of this finding should be qualified somewhat since the pretest score had a low relationship with the posttest ($r = .34$, $p < .05$), suggesting it was a somewhat poorer predictor of outcomes than prior knowledge usually is (Dochy and Alexander, 1995). Therefore, conclusions about the relative importance of prior knowledge and monitoring accuracy will have to await research in which the pretests specifically sample students' prior knowledge of the instructional domain, rather than general knowledge of that domain, or when instructional efficacy can be manipulated experimentally.

Finally, since the reading comprehension test score did not contribute significantly to the prediction of posttest scores while general knowledge monitoring did, the results also suggest that students' reading comprehension was not as important in predicting learning as was their general monitoring accuracy. Again, some caution in generalizing these results is in order for several reasons. The reading comprehension test score and the KMA were highly correlated ($r = .81$), suggesting colinearity between the two measures. This is not too surprising because the words for the general monitoring KMA were selected from prior versions of a reading achievement test, and one would expect such materials to be highly correlated. The correlation of the reading comprehension score and the learning theory KMA while significant ($r = .42$, $p = .01$), was much lower. Finally, the distribution of reading comprehension scores was positively skewed, with most of the scores falling at the upper end of the distribution. Thus, further research with reading materials and vocabulary words which are not as closely related is needed to examine the relative contributions of knowledge monitoring to learning.

Other Relevant Studies

Several studies reported earlier dealt with the issue of the domain generality–specificity of metacognitive knowledge monitoring. Study V, for example, examined the domain generality–specificity issue by exploring the relationship of vocabulary knowledge monitoring assessments to different types of instruction in specific domains, sonic transmission in underwater environments. The KMAs used common words of varying difficulty, and less common, technical words dealing with oceanography. The oceanography words were similar to the instructional domain without duplicating it. The domain generality question was investigated in two ways: (1) by comparing the correlations between knowledge monitoring accuracy in the two content domains, and (2) by examining the relationship of each word set with success in a training course. The correlations between the two KMAs on the first administration was .44 ($p < .001$). That relationship increased to .59 ($p < .001$) for the second administration of the oceanography words after the trainees had read the instructional text. These data suggest a moderate relationship between monitoring accuracy in the different domains.

The two KMAs were also correlated with trainees' course learning. These correlations are shown in Table 6. When more common, general vocabulary was used, the

Table 6

Correlations Between KMA and Learning in Training Course

KMA Score Coefficients for Various Word Types	*r*
Easy (13 Words)	.24*
Medium Difficulty (14 Words)	-.01
Difficult (13 Words)	.36**
Common Oceanography (15 Words)	.24*
Technical Oceanography (25 Words)	-.01
Common Oceanography–Administration 2	.39**
Technical Oceanography–Administration 2	.33**
40 Popular Words	.31**
40 Oceanography Words	.16
40 Oceanography Words–Administration 2	.45**

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

KMA was related to learning outcomes only for the easy and difficult words. Upon initial testing, the KMA was correlated with the more common oceanography words but not the more technical words. After the instructional text was read, the KMA correlations increased for both the common and technical oceanographic terms.

Students may have been able to recall or infer the meanings of many technical words from the text, and the higher correlations on the second administration of the KMA are a further indication that knowledge monitoring accuracy is important for learning. In sum, the correlations of the different KMA word categories with learning outcomes indicate that knowledge monitoring accuracy was related to learning in a technical domain, again suggesting moderate generality for knowledge monitoring accuracy. In this study, and in those reported by Schraw et al. (1995), the evidence regarding domain generality of metacognition is mixed.

Returning to Study VI, two different KMA scores, verbal and math, were reported. The correlation between the two KMA scores was .33 ($p < .001$) indicating a low but significant relationship between knowledge monitoring in those two domains. The math KMA score correlated .52 ($p < .001$) with the SAT I mathematical scores, but was not related to the SAT I verbal scores. The verbal KMA had correlations of .27 ($p < .001$) with the SAT I–V, and .16 ($p < .001$) with scores on the SAT I–M, apparently reflecting some of the verbal content of the math word problems. Some students (n=93) had taken the SAT I a second time. The math KMA–SAT I–M correlations for this subgroup was .56 ($p < .001$), and correlations between verbal KMAs and SAT I–V were .44 ($p < .001$), and .28 with the SAT I–M ($p < .01$). As in the Schraw et al. (1995) study and in the research described earlier, these results suggest both domain specific and domain general components to knowledge monitoring. The higher relationships with SAT scores in the same content domain point to domain specificity, and the significant correlations between the verbal and math KMAs, as well as the significant, though low, correlations of the verbal analogies KMA with the SAT I–M hint at domain generality.

The data dealing with the generality–specificity issue indicate that there are both domain general and domain specific components to knowledge monitoring. While the results vary among studies, correlations between KMAs from different domains, and between KMAs and learning in different subjects are generally in the low to moderate range and never approach levels suggesting that knowledge monitoring is either predominantly domain general or domain specific. Much of the content taught in schools is similar across the curriculum. This complicates attempts to study the generality–specificity of knowledge monitoring, since correlations with knowledge monitoring in different fields not only reflect relationships between knowledge monitoring accuracy, but also some of the relationships between different content domains. A definitive study of this issue may require use of a domain which is entirely unfamiliar to the participants, development of a KMA in that domain, and then comparing relationships between novel and more familiar domains. Pending such research, results are likely to indicate that metacognitive knowledge monitoring has both domain general and domain specific components.

VI. Self-Reports of Metacognition and Objective Measures of Knowledge Monitoring

In addition to the work on metacognition there has been substantial interest in what the learner is thinking about prior to, during, and after learning activities and the role that these cognitive processes play in facilitating learning in complex environments (i.e., the strategic aspects of learning). Weinstein refers to these as learning and study strategies, which include "a variety of cognitive processes and behavioral skills designed to enhance learning effectiveness and efficiency" (Weinstein and Mayer, 1986, p. 4). As mentioned above, a number of such self-report measures of metacognition and self-regulated learning are available (Jacobs and Paris, 1987; Pintrich, Smith, Garcia, and McKeachie, 1991; Schraw and Denison, 1994; Tobias, Hartman, Everson, and Gourgey, 1991) and the relationships between these and the KMA were investigated in two studies reported below.

Self-reports of metacognitive processes are subject to a number of problems which have been described elsewhere (Tobias and Everson, 2000) and will be summarized here only briefly. Since metacognition involves monitoring, evaluating, and coordinating cognitive processes, many have wondered if students are aware of the processes used during learning? Can they describe them by merely selecting from alternatives on a multiple choice scale? There is also a question about whether students are reporting honestly. While the truthfulness of self-report responses is always an issue, it is especially troublesome for metacognitive assessments because students may be reluctant to admit exerting little effort on schoolwork, especially when those reports may be available to their instructors.

The KMA also uses self-reports, the estimates of knowledge or ability to solve problems, in addition to students' test performance. However, unlike questionnaires about students' metacognition the KMA does not ask participants to report on cognitive processes used while performing a task. Estimates on the KMA are more accessible than reports of cognitive processes and, therefore, the KMA is less likely to be affected by the difficulties of abstract recall involved in responding to metacognitive questionnaires. This was confirmed by research (Gerrity and Tobias, 1996), suggesting that the KMA, compared to self-report scales of test anxiety, was less susceptible to students' tendency to present themselves in a favorable light. Despite these differences, a modest relationship between metacognitive self-reports and the KMA was expected. These issues were addressed in Study VI and VIII and will be reviewed next.

In addition to investigating the KMA's relationship with ability in Study VI, we examined the correlations between the KMA and three other assessments of students' metacognition—two self-report questionnaires and teachers' observational reports. In view of their obvious content and methodological differences, moderate relationships between the KMA and these other assessments was expected. In addition to the verbal and math KMAs, participants completed two self-report scales: (1) the Learning and Study Skills Inventory (LASSI) (Weinstein et al., 1987); and (2) the Metacognitive Awareness Inventory (Schraw and Dennison, 1994)—a 51-item self-report scale measuring different aspects of students' metacognitive processes. Moreover, teachers assessed students' metacognitive abilities by completing a seven-item, Likert-type scale.

Students' LASSI scores were factor analyzed, and two factors accounting for 68 percent of the variance were extracted. The first factor described students' reports of relatively automatic study strategies. A representative item was *...I am distracted from my studies very easily*. Items loading on the second factor reflected the use of strategies requiring deliberate effort, for example *...I use the chapter headings as a guide to find important ideas in my reading*. Their LASSI factor scores were correlated with their KMA math and verbal scores, with correlation of $r=.21$ between LASSI Factor 1 and the math KMA and $r=.19$ with the verbal KMA. The correlations with Factor 2 were $r=-.11$ and -.20, respectively. While these correlations were, with one exception, statistically significant because of the large number of cases, the absolute relationships were small and somewhat lower than expected. One reason for the low relationships may be that the LASSI does not have a scale designed to measure metacognition, in general, or knowledge monitoring, in particular.

The MAI scores were also factor analyzed and yielded a single factor accounting for 69 percent of the variance. This factor was correlated only with the math KMA ($r=.12$, $p < .05$). The seven-item teacher rating scale was also factor analyzed, yielding two factors with Eigen values above unity and accounting for 71 percent of the variance. Varimax rotation results were then submitted to a second order factor analysis yielding one monitoring factor which correlated significantly only with the math KMA ($r=.143$, $p <.05$).

Not surprisingly, the most striking aspect of these results was the relatively high correlations among the self-report measures, as indicated in Table 7, when contrasted with correlations with the KMAs. We suspect that the self-report measures are sharing common method variance.

The correlations among the KMAs and both the LASSI and MAI are sometimes significant, but generally quite low, suggesting relatively small amounts of shared construct variance. Moreover, the higher KMA correlations with the SAT, compared with the SAT I correlations with the LASSI or MAI, may be attributable to the fact that the content of the KMAs was adapted from the SAT I and SAT II. Also, in prior research (Everson, Tobias, and Laitusis, 1997), the KMA was often related more to GPA than the LASSI. Thus, a strong relationship between the KMA and scholastic aptitudes was not unexpected.

In Table 7 the teacher rating scale had low relationships with all of the other variables. Those results may be attributed to the fact that many of the teachers refused to fill out the rating form, indicating that such an assessment was foreign to the philosophy of the

Table 7

Correlations of Verbal and Math KMA Scores, Metacognitive Self-Report Scales, Teacher Ratings, and SAT I–V and SAT I–M Scores

Variables	*LASSI-I*	*LASSI-II*	*MAI*	*Teacher Rating*	*SAT I–M*	*SAT I–V*
Verbal KMA	.19**	-.20***	.01	.06	-.01	.29***
Math KMA	.21***	-.11*	.12*	.16*	.50***	.16*
LASSI Factor I	—	—	.25***	.18	.06	.08
LASSI Factor II	—	—	.56***	.06*	-.10*	-.10*
MAI Factor	.25***	.56***	—	.11	.01	-.11*
Teacher Rating Factor	.19*	.06	.11	—	-.08	-.00

* $p < .05$
** $p < .01$
*** $p < .001$

educational program in which they were engaged. Furthermore, we noted that the ratings had limited variability—the teachers were reluctant, apparently, to rate these gifted and talented students low on any academically related measure.

In addition to the lack of shared method variance, it is plausible that the low correlations between the KMA and LASSI factors in Study VI were due to the absence of a LASSI scale designed specifically to assess metacognition. In an effort to examine this possibility, another self-report instrument with a metacognition subscale, the Motivated Strategies Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia, and McKeachie, 1991), was used in Study VIII. Unlike the LASSI, which assesses use of study strategies in general, students respond to the MSLQ in terms of the strategies used in particular courses. Therefore, in Study VIII we examined the relationships between the KMA and the MSLQ, the relationships between these assessments and the LASSI, as well as students' GPA which were discussed earlier.

Recall that in Study VIII participants (n = 120) were given math and verbal KMAs, the LASSI, and the MSLQ. Participants' GPAs were retrieved from the college's records. The MSLQ was administered using standard instructions requiring students to respond in terms of a particular course. Half the sample was asked to respond in terms of an English course they were currently taking, while the other half responded in terms of a mathematics course. If students were not registered for courses in either content area, they were asked to respond in terms of the last course of that type they had taken.

The 10 LASSI subscales were factor analyzed and two interpretable factors, similar to those found earlier, emerged. One was related to automatic study strategies and the other was related to strategies under volitional control. Since the MSLQ was administered to half the participants in terms of an English class and the other half in terms of a mathematics course, the number of students completing either version of the MSLQ was only about half the total sample size, and too small for factor analysis. The correlations between the MSLQ scales, including the one assessing metacognition, and the math or verbal KMAs, were quite low. However, in order to determine whether all the MSLQ scales assessed a construct similar to the KMA, multiple correlations were computed for students' scores on all 15 MSLQ subscales and their math and verbal KMA scores. Separate correlations were computed for students completing the scales in terms of English or mathematics courses. These data are shown in Table 8.

The multiple correlations between the MSLQ and the KMAs were substantially higher than the zero order correlations, though often failing to reach statistical signifi-

Table 8

Correlations of MSLQ, LASSI, KMA Scores, and GPA

Variables	*Math KMA*	*Verbal KMA*	*LASSI Factor I*	*LASSI Factor II*	*Total GPA*
MSLQ-E1	.68	.61	.66	.90**	.45
MSLQ-M1	.55	.40	.77**	.82**	.54
LASSI-I	-.03	.11	—	—	.32**
LASSI-II	-.20*	-.17	—	—	.14
Math and Verbal KMA	—	—	.15	.22	.36**

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

cance. Clearly, further research with larger samples is needed to determine the MSLQ's relationship with the KMA. The correlations between the math and verbal KMAs and the LASSI factors were similar to those reported in Study VI, though they were not statistically significant because of the small number of subjects. Surprisingly, when the ten LASSI subscales were regressed onto the math KMA scores the multiple correlation was $r=.46$, and the multiple correlation with verbal KMA score was $r=.35$, essentially doubling the multiple correlation with LASSI factor scores. This discrepancy in results is attributable, we believe, to the fact that multiple correlations tend to maximize relationships among predictors and criteria, while the factors tend to optimize an interpretable factor structure. Thus, these somewhat higher multiple correlations may be misleading.

All these assessments were related to students' academic achievement as measured by undergraduate GPA. These data, also displayed in Table 8, indicate that the MSLQ had moderate relationships with achievement, though again failing to reach statistical significance due to the small number of cases and multiple predictors. The correlation between the verbal and math KMAs were $r=.49$ ($p < .001$).

The results suggest a convergence between the overall MSLQ and knowledge monitoring, though replication is needed because the correlations, while of reasonable magnitude, failed to reach statistical significance because of the small sample and the large number of predictor variables. Correlations among the KMA scores and the LASSI factors were also low, though the multiple correlations improved that picture somewhat. It seems reasonable to conclude that the MSLQ and the LASSI, in total, are measuring constructs similar to the KMA, though none of the individual self-report scales are similar in content to the KMA. The correlations in Table 8 also indicate that the LASSI and MSLQ were markedly similar (multiple *r*'s ranging from .7 to .9), not surprising since many of the items in both scales are quite similar and both assessments share method variance.

Summary: Relationships of KMA Scores and Self-Report Measures of Metacognition

As expected, low-to-moderate relationships were found between the different self-report instruments used in the various studies and the KMA. Several findings were consistent over the different studies. The self-report questionnaires tended to correlate highly with one another, indicating that they measured similar characteristics using quite similar methods. The uniformly lower correlations with the KMAs suggest that though the KMAs evaluate some of the same characteristics as the self-report measures, the differences in method may diminish their shared variance.

In most of the studies the correlations among the KMAs and measures of scholastic aptitude and learning tended to be higher than those for the self-report instruments. This suggests the variance that the KMA does not share with the self-report instruments is more highly associated with school learning than are the questionnaire measures. A recent study, conducted in Belgium (Van Damme and Claesen, personal communication) found low, and often insignificant correlations, between a vocabulary based KMA and several self-report measures. Especially interesting were findings that the KMA scores were strongly related to teachers' judgements of metacognition, while the self-report scales were not. This research confirms our earlier findings that, in contrast to self-report measures, the KMA assesses components of metacognition that are more highly related to classroom learning.

Study X: An Investigation of the Impact of Anxiety on Knowledge Monitoring[11]

The task of knowledge monitoring is complex. Skilled metacognitive knowledge monitoring requires that students retrieve prior declarative or procedural knowledge from long-term memory, compare it to material with which they are presented, and make a judgment as to whether they know the material, or are able to use it to solve problems. The complexity of these tasks suggests that knowledge monitoring should be negatively affected by anxiety. Anxiety is an unpleasant affective state, similar to fear, that has been shown to debilitate performance on complex cognitive tasks (Hembree, 1988; 1990). It has been suggested (Tobias, 1992) that anxiety impairs performance because the central representation of anxiety absorbs cognitive resources, leaving reduced capacity for students to deal with demanding tasks.

Prior research (Tobias and Everson, 2000) found negative relationships between anxiety and knowledge monitoring. Those studies assessed students' anxiety by self-

[11] The description of this study is based on a dissertation completed by Julie Nathan at Fordham University in May 1999, and on a report of that study by Nathan and Tobias (2000) presented at a meeting. Both sources may be found in the references.

report inventories. Such inventories usually pose a variety of problems, as indicated in the section dealing with self-reports below. Thus, one purpose of this study was to manipulate anxiety experimentally in order to make more causal inferences about the impact of anxiety on knowledge monitoring. A further purpose of this study was to examine the impact of anxiety at different points during the KMA procedure, i.e., during estimation, testing, or help seeking, in order to determine when anxiety induces its most debilitating effects. It was expected that clarification of that question would provide a better understanding of how anxiety exerts its effects.

Participants and Procedures

Participants (n=97) were drawn from a small parochial high school in a rural community. Students completed a number of anxiety measures and were assigned randomly to one of four experimental groups. The Mathematics Anxiety Scale (MAS), a subscale of the Fennema-Sherman Mathematics Attitude Scales (1976) was used to assess math anxiety. Students reported their level of state worry, i.e., the degree of anxiety experienced at that time, on an abbreviated version of the Worry-Emotionality Questionnaire (WEQ) (Liebert and Morris, 1967; Morris, Davis and Hutchings, 1981) consisting of only five worry items. Participants were directed to report their feelings and attitudes in reaction to the tasks presented to them.

The KMA consisted of 26 math problems adapted from prior versions of the SAT. In a second session, one week later, all groups performed the same four tasks. The experimental manipulation consisted of varying the point at which evaluative stress was induced. Stress was induced by telling students the KMA was a measure of ability requiring much effort. The four groups were: (1) an estimation group in which students estimated whether they could or could not solve each of 26 mathematics problems and received stress-inducing information at this point; (2) a test group that took a multiple-choice test containing the same KMA problems and stress instruction were given at this point; (3) a help group which was permitted to select any test items on which they wished to receive additional help (answers to the math problems and solution strategies), again, stress instructions were administered at this point; and (4) a control group which received no stress-inducing instructions. Immediately after the experimental sessions, participants completed another brief Worry scale to determine the level of test anxiety experienced at that point (see Nathan, 1999 for additional details).

Results and Discussion

There were significant differences in students' reported worry, indicating that the stress instructions succeeded in arousing varying levels of anxiety across the four experimental groups. It was expected that anxiety would peak when evaluative stress was introduced for each group but contrary to expectations results showed that the anxiety increased similarly for all groups over the different tasks (F [2, 192]=29.20, $p < .01$). from a low of 9.59 at estimation to 12.8 at help seeking. Apparently, once stress was introduced it continued at an elevated level throughout the experiment; this was true even for the control group which received no stress-inducing instructions. Presumably, the anxiety usually connected with solving math problems and the test-like context were sufficient to arouse and maintain test anxiety throughout the experiment. Anecdotal evidence indicated that students knew quite a bit about the task even before the instructions were given, suggesting that prior participants told succeeding ones about the tasks and that this feedback mitigated the influence of the anxiety-inducing instructions.

Consistent with prior findings, those high in math anxiety performed more poorly on the math test (r=-.36) and were less accurate knowledge monitors (r=-.21) than those lower in anxiety (Tobias and Everson, 2000). They also estimated being able to solve fewer problems than their less anxious peers (r=-.49). However, contrary to expectations, there were no differences among the groups on any of these dependent variables, presumably because of the feedback among groups. One important suggestion for future research is that participants assigned to different groups should either be examined in different places making the possibility of such feedback unlikely, or if they are recruited from the same location, they should be examined at the same time. Such precautions are important in situations where differential instructions are given.

Study XI: Cross-Cultural Perspective on Knowledge Monitoring[12]

All prior KMA studies were conducted in the United States using American students. In Study XI we examined knowledge monitoring in other cultures by comparing Nigerian and American students. Nigerian students were of special interest since there is some evidence suggesting that instruction in that country relies heavily on rote memorization (Kozulin, 1998). It was, therefore, interesting to

[12] The data for this study were collected by Hyacinth Njoku, Graduate School of Education, Fordham University.

determine whether such instructional practices led to less accurate metacognitive knowledge monitoring and less strategic help seeking among Nigerian students.

Participants and Procedures

The participants were tenth- and eleventh-grade high school students from the United States and Nigeria. The Nigerian sample (n = 77) was composed of male students from an elite Catholic High School. The American sample consisted of two groups from two different parochial high schools. One group (n = 70, 39 females) came from a Catholic coeducational high school in an urban middle class neighborhood in a small city in northeastern United States. The second group (n = 57) was all male, predominantly minority students from lower socioeconomic backgrounds attending a Catholic high school in a poor urban neighborhood in the Bronx, New York.

All students received a KMA composed of 24 math problems selected by experts as being equally appropriate for Nigerian and American students. They first estimated their ability to solve each of the problems, then took a multiple-choice test on the problems. Finally, they were given the option of receiving additional help on any of the problems. The help consisted of the correct answer to the problem and the steps to reach the answer.

Results and Discussion

Math KMAs were computed for all students and the data indicated substantial differences between the two groups of American students. Therefore, they were analyzed as separate groups. A one-way ANOVA among the four groups (Nigerian males, and the three American groups of females, middle class males, and males from poorer backgrounds) found substantial differences in knowledge monitoring accuracy (F [3,200] = 4.42, $p < .005$). Post hoc contrasts revealed that middle class participants were more accurate monitors than the minority group ($p < .004$), and that the majority American males were marginally more accurate than the Nigerian students ($p < .056$). There were no differences in monitoring accuracy between Nigerian students and minority American males or females.

A one-way ANOVA of the math achievement raw scores revealed a significant main effect on math performance (F [3,200] = 12.361, $p < .000$). Post hoc contrasts indicated that middle class American males had higher scores than Nigerians and poorer American males. Math achievement scores of poorer American males were lower than those of the Nigerian or American female students. Thus, while the Nigerian males differed only marginally in knowledge monitoring accuracy from their American counterparts, the differences in math achievement between these groups was more substantial. These data suggest that poor performance in mathematics among Nigerian students is not related to poor metacognitive knowledge monitoring. Clearly, cross-cultural differences in metacognitive knowledge monitoring need further study with larger and more comparable groups of students.

VII. Metacognitive Knowledge Monitoring and Strategic Studying Among Secondary Students

There is a compelling rationale for the relationship between knowledge monitoring and strategic studying, especially at secondary and postsecondary levels where a great deal of instructional material has to be mastered in a short time. Students who accurately differentiate between what they know and do not know are likely to study more strategically. Furthermore, as suggested by the results of Study II and III, it was also expected that students with accurate metacognitive knowledge monitoring skills are apt to seek help on tasks they estimated as knowing yet failed on a subsequent test. These expectations were examined in Studies X and XI.

As described earlier, in Study X a 26-item math KMA was administered along with a number of anxiety measures designed to assess both math anxiety and the degree to which anxiety was experienced while working on the experimental tasks. Students were assigned randomly to one of four groups that varied as to when they received stress-inducing instructions. Strategic review was studied in two ways. Responses were scored by assigning one point for each item that students failed on the test and on which they subsequently sought help and one point for those items that students passed on the test on which help was not requested later. The results indicated that more anxious students sought help on a significantly greater percentage of items passed on the test, a nonstrategic and inefficient endeavor. The group receiving stress instructions during the test reviewed significantly more items than the other three groups, perhaps because the experimental situation for them was most similar to stressful school examinations.

Similarly, we expected accurate knowledge monitors would review more strategically (i.e., would omit reviewing material they had mastered already) than less accurate students. Results from the control group, who received no stress instructions, would most directly address this question. Knowledge monitoring accuracy and strategic review were positively correlated in the control group, but this relationship did not reach statistical significance. Monitoring accuracy and strategic review were expected to correlate negatively in the three groups receiving stress instructions because the evaluative stress instructions were expected to disrupt the otherwise positive relationship between these variables. Contrary to expectations, however, only the data from the *Test group* showed a negative relationship between these variables. This finding suggests that accurate knowledge monitors who feel most stressed just before taking an examination were least strategic in deciding on which items to ask for additional assistance after the exam.

In the second analysis, a repeated measure ANOVA was computed on the types of items on which participants sought help. Consistent with Studies II and III, the data were analyzed in terms of the percentage of items on which help was sought from each of four categories. Items students estimated knowing that were later (1) passed (+ +) or (2) failed (+ -) on the test, and words students estimated not knowing which were later (3) passed (- +) or (4) failed (- -) on the test. These data are shown in Figure 5. The differences were statistically significant, indicating that the more anxious students reviewed a greater percentage of items passed on the test, clearly a nonstrategic and inefficient endeavor.

The repeated measures ANOVA also revealed that the *Test group's* item review pattern was least efficient. Although they reviewed a greater proportion of items which they failed on the test, they reviewed many items

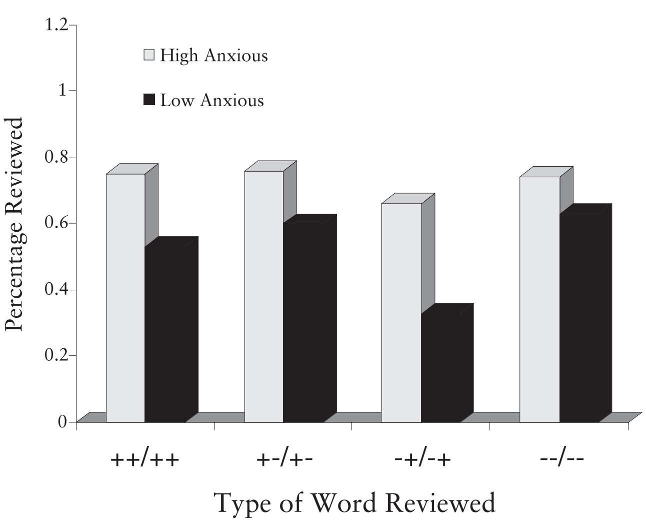

Figure 5. Relationship between test anxiety and help seeking.

they got right on the test. These review patterns suggest that feeling stressed before the test interferes with the ability to review efficiently and strategically. An alternative explanation may be that the *Test group* reviewed more items than the other three groups, perhaps because that condition was most similar to stressful school exams.

In Study XI we also examined strategic help seeking. Again there were four groups of 10th and 11th Nigerian and American high school students: (1) 77 Nigerian male students from an elite Catholic high school; (2) 70 male and (3) 39 female students from am American Catholic coeducational high school in a middle-class neighborhood; and (4) 57 minority students from low socioeconomic backgrounds attending a Catholic high school in a poor urban community. The KMA had 24 math problems. After estimating their ability to solve each problem, students took a multiple-choice test comprised of these same problems. Finally, they were given the option of receiving additional help on any of the problems. The help consisted of seeing the correct answer to the problem and the steps required to solve the problem.

The data were analyzed using a 2 (high/low KMA) X 4 (++, +-, -+, --) ANOVA, with the four types of items treated as a repeated measure. An overall multivariate difference among the items was found (Wilks' Lambda = .875, F [3, 195]=, 9.25, $p < .001$), but there was no difference between low and high monitoring groups, nor was there an interaction between monitoring groups and type of items reviewed. A similar analysis examining differences among cultural group was found (F [3, 197] = 2.60 < .054). Post hoc contrasts revealed that the differences were for those items students estimated not knowing and failing (--) with American middle-class students seeking more help on these items than Nigerian students. These data indicate somewhat more strategic performance by the middle-class American males, though a subsequent comparison revealed differences ($p > .027$) in items estimated to be unknown and passed on the test (-+), with Nigerian males reviewing fewer of these items than the middle class sample of American males. The means for the different items by group are displayed in Figure 6.

The mixed findings regarding help seeking and the small sample sizes indicate that the Nigerian–American cultural differences in metacognitive behaviors require re-examination with larger samples.

Moreover, there were contradictory findings regarding strategic help between Studies II and III and X and XI. The first two studies confirmed that those who accurately differentiated between what they know and do not know sought help more strategically by omitting known material and seeking help for unfamiliar material. In contrast, less accurate monitors generally

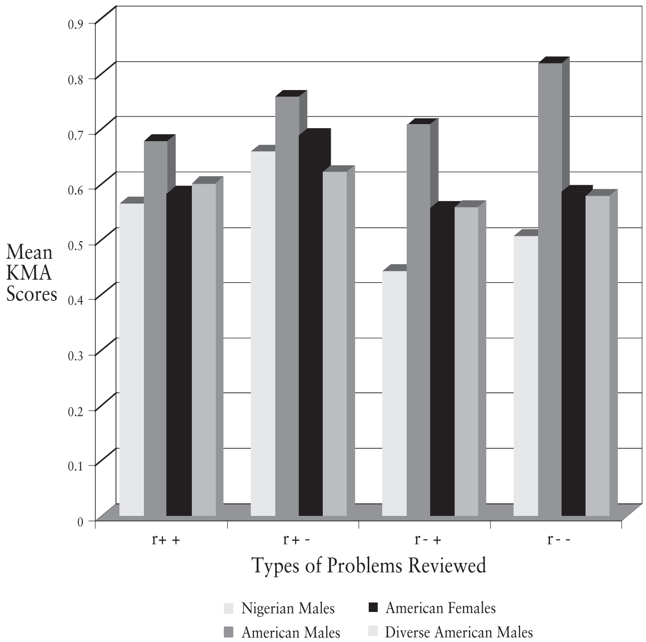

Figure 6. Math KMA means by type of problem reviewed for Nigerian and American male students.

allocated their attention less effectively by studying what was already known at the expense of unfamiliar material. In Studies X and XI, however, accurate knowledge monitoring had little impact on strategic help seeking. It is plausible that these contradictory findings are due to sampling and methodological differences in the research designs. The first two studies used elementary school students, while secondary school students participated in the later studies.

Secondary school students, and those at postsecondary levels, have to master a great deal of new material. In that setting, the time needed to seek help has important implications since students usually do not have enough time to get help on all the items on which they might want assistance. Therefore, help should be sought only on items where it is most needed. In the conditions implemented in the later studies, however, students could receive help on all the items they wished, hence they had no need to act strategically. An important suggestion for further research is to permit help on a small number of items so assistance on some items is obtained at the expense of others. Ideally students should seek help only on those they estimated knowing/being able to solve and subsequently failing on a test, as well as items estimated as unknown/unable to solve and failed on test. Reviews of the results of several studies using vocabulary, verbal analogies, and math KMAs indicate that approximately 36 percent of students' responses fall into those categories. Therefore, one suggestion might be to permit students to obtain help only on a portion of the presented items. Then, the help condition may more nearly reflect the learning conditions students encounter in schools.

There are two other suggestions for further research. First, all our studies used relatively small samples, thus replications and extensions with larger numbers of participants is clearly needed. Second, it would be useful to obtain "think aloud" protocols as students engage in help seeking or review. Such protocols may clarify students' thinking about their motivations for help seeking on different types of items.

VIII. General Discussion

The 11 studies presented here, along with our earlier research, lend additional support to the validity of the KMA. In sum, they demonstrate the importance of assessing knowledge monitoring ability as a way to enlarge our understanding of human learning. For the most part, the findings bolster the importance of accurate knowledge monitoring for school learning, reading at elementary school levels, strategic learning at least among elementary school students, and the development of academic aptitude. The studies extended our earlier work on the KMA in two major ways: (1) the content of the KMA was extended to mathematics at secondary and postsecondary school levels, as well as to verbal analogies and science, and (2) new populations were studied, including the gifted and talented, adults in military training programs, and foreign-born students.

The research dealing with general ability provided evidence that knowledge monitoring was positively related to general ability. Prior findings relating the KMA to academic achievement implied that knowledge monitoring was related to how much students knew. Since achievement and aptitude are highly related, positive relationships with ability were expected, and these studies provided empirical evidence of that relationship. Results varied with respect to the size of the ability–monitoring relationship, and further research is needed to examine the common variance between the two constructs. Nonetheless, we expect a positive relationship between these cognitive components. At the same time we expect that knowledge monitoring will have some unique variance, since the ability to estimate one's knowledge should be related to how much is known, but not be synonymous with it.

This body of research suggests that the ability to differentiate between what is known (learned) and unknown (unlearned) is an important ingredient for success in all academic settings. It should be noted, however, that accurate monitoring of one's prior knowledge is only one aspect of metacognition that may be important for success in school-like situations. Good students also have to monitor their comprehension of reading assignments and their understanding of lectures and other knowledge acquisition efforts. In addition, good students typically consult other resources (students, teachers, or reference works) in order to enlarge their understanding, answer questions, and integrate material learned in different courses. They also need to check whether they have completed assignments in different courses, whether they have met the various requirements in order to progress in their academic careers, etc. These are only a few of the different types of monitoring behaviors students need to perform in order to succeed in school. These various forms of metacognitive monitoring have been studied as part of the research on academic self-regulation (Winne, 1996; Zimmerman and Risenberg, 1997).

Although we are aware of no research specifically relating monitoring of prior knowledge and academic self-regulation, there is a strong basis for expecting such a relationship. The ability to differentiate what one knows from what is still to be mastered is likely to be a fundamental component of other forms of self-monitoring and self-regulation. It is hard to imagine successful monitoring of reading comprehension or understanding lectures if students cannot accurately differentiate between what they know and do not know. It, therefore, may be hypothesized that accurate monitoring of prior knowledge is a prerequisite for the effective self-regulation of learning, as we assume in our model (see Figure 1). Clearly, research examining these relationships is urgently needed.

Self-regulation is a compound construct made up of cognitive, emotional, motivational, and conative (students' willingness to expend effort to accomplish their goals) components; these are sometimes referred to as "skill and will" components (Weinstein and Mayer, 1986; Zimmerman and Reisenberg, 1997). Thus, students may have the cognitive abilities needed for effective self-regulation but lack the motivation, desire, or will to regulate their efforts. Therefore, students who are accurate knowledge monitors may not be effective at regulating their academic behavior, unless they also have the motivation, energy, and the will to do so.

IX. Suggestions for Future Research

In most of the KMA studies, the majority of students' estimates (47 to 69 percent in a sample of studies) are that they know or can solve a particular item. Thus, even though the Hamann coefficient is a measure of the accuracy of students' estimates, actual knowledge and accuracy are slightly confounded. It would be useful to have a greater number of estimates in the "do not know/cannot solve" category. One way of achieving this is to add items consisting of nonsense words or insoluble problems. When confronted with such items or tasks, students, especially those who are quite knowledgeable, will have to estimate not knowing the item, thereby helping to disentangle the possible confounding between knowledge monitoring and estimation.

Further research on the relationship between knowledge monitoring, motivation, and conation is clearly needed to determine their impact on students' ability to be strategic learners. It would be useful to identify students who are accurate knowledge monitors, but not highly motivated to succeed academically. We would expect that, despite their knowledge monitoring skills, such students will not effectively control their studying behaviors and, therefore, not do well academically. On the other hand, highly motivated students who are willing to invest considerable effort to attain academic goals are unlikely to be effective unless they also possess accurate knowledge monitoring skills. Knowledge monitoring in general, and the KMA in particular, is less likely to be affected by motivational and volitional variables than are more complex self-regulatory behaviors, since making knowledge or problem solving estimates is relatively automatic and takes little effort. Therefore, motivation and conation should have less effect on such judgments. On the other hand, obtaining additional help and/or reviewing prior learning does require more effort and reflection than making estimates and may, therefore, be more affected by motivational and volitional tendencies.

It may also be helpful to obtain "think aloud" protocols during the estimation phase of the KMA procedure. Such protocols can clarify students' thinking when they are estimating their knowledge of items. It would be especially interesting to determine whether they have any doubts while estimating knowing items that are subsequently failed. It may be expected that accurate monitors are more likely to have some doubts about these items than their less accurate peers.

Logical analysis has led us to assume a hierarchical organization of metacognitive activities. That is, we presumed that such more advanced metacognitive activities as evaluating learning, selecting appropriate learning strategies, or planning future learning activities could not occur without accurate knowledge monitoring. While the logic seems compelling, it should nevertheless be subjected to empirical test. It would be interesting to observe students in classes while they conduct self-evaluations of their work, select strategies, or plan for future learning and determine whether accurate monitors do in fact conduct these activities more efficiently.

Finally, it has been assumed that the KMA is more resistant to the problems posed for self-report instruments in general, such as students' tendency to make socially desirable responses. That assumption should be studied experimentally, by instructing some students to respond to the KMA in ways to make themselves seem to be very diligent students, while others are instructed to appear casual or lackadaisical; of course, a control group should receive neutral instructions. In such studies, participants should also be asked to respond to self-report measures to study the relative resistance of both types of assessment to student "faking."

References

Borkowski, J.G. (2000, April). *The assessment of executive functioning*. Paper delivered at the annual convention of the American Educational Research Association, New Orleans.

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, NJ: Lawrence Erlbaum Associates.

Butler, D.L. & Winne, P.H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245–281.

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

Dochy, F.J.R.C. & Alexander, P.A. (1995). Mapping prior knowledge: A framework for discussion among researchers, *European Journal of Psychology of Education, 10*, 225–242.

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

Everson, H.T. & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science, 26*, 65–79.

Everson, H.T., Tobias, S., & Laitusis, V. (1997, March). *Do metacognitive skills and learning strategies transfer across domains?* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Fennema, E. & Sherman, J.A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes towards the learning of mathematics by females and males. *Catalog of Selected Documents in Psychology, 6*, 3, 1–32.

Fajar, L., Santos, K., & Tobias, S. (1996, October). *Knowledge monitoring among bilingual students*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.

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

Gates, W.H. & MacGinitie, R.K. (1989). Gates MacGinitie Reading Test (3rd Ed.). Chicago: The Riverside Publishing Co.

Gerrity, H. & Tobias, S. (1996, October). *Test anxiety and metacognitive knowledge monitoring among high school dropouts*. Paper presented at the annual convention of the Northeastern Educational Research Association, Ellenville, NY.

Glenberg, A.M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 166*, 119–136.

Goodman, L.A. & Kruskal, W.H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732-764.

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

Guthke, J. (1992). Learning tests: The concept, main research findings, problems, and trends. *Learning and Individual Differences, 4*, 137–151.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*, 33–46.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*, 47–78.

Jacobs, J.F. & Paris, S.G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist, 22*, 255–278.

Jimenez, R.T., Garcia, G.E., & Pearson, P.D. (1995). Three children, two languages, and strategic reading: Case studies in bilingual/monolingual reading. *American Educational Research Journal, 32*, 67–97.

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

Kouzlin, A. (1998). *Psychological Tools: A sociocultural approach to education*. Cambridge, MA: Harvard University Press.

Liebert, R.M. & Morris, L.W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*, 975–978.

Lidz, C.S. (1992). The extent of incorporation of dynamic assessment into cognitive assessment. *Journal of Special Education, 26*, 325–331.

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.

Nathan, J. (1999). *The impact of test anxiety on metacognitive knowledge monitoring*. Unpublished doctoral dissertation completed at Fordham University, NY.

Nathan, J. & Tobias, S. (2000, August). *Metacognitive knowledge monitoring: Impact of anxiety*. Paper delivered at the annual meeting of the American Psychological Association, Washington, DC.

Nelson, T.O. (1984). A comparison of current measures of the accuracy of feeling of knowing predictions. *Psychological Bulletin, 95*, 109–133.

Nelson, T.O. & Nahrens, L. (1990). Metamemory: A theoretical framework and new findings. In G.H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). New York: Academic.

Pintrich, P. R., Wolters, C.A., & Baxter, G.P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw (Ed.) *Issues in the measurement of metacognition*. Lincoln, NE: Buros Institute/ The University of Nebraska.

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.

Romero, R. & Tobias, S. (1996, October). *Knowledge Monitoring and Strategic Study*. Paper presented at a symposium on “Metacognitive Knowledge Monitoring” at the annual convention of the Northeastern Educational Research Association, Ellenville, NY.

Romesburg, H.C. (1984). *Cluster analysis for researchers*. London: Wadsworth, Inc.

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

Schommer, M. & Walker, K. (1995). Are epistemological beliefs similar across domains? *Journal of Educational Psychology, 87*, 424–432.

Schraw, G. (2000). *Issues in the measurement of metacognition*. Lincoln NE: Buros Institute of Mental Measurements and Erlbaum Associates.

Schraw, G. (1995). Measures of feeling of knowing accuracy: A new look at an old problem. *Applied Cognitive Psychology, 9*, 329–332.

Schraw, G. & Denison, R.S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460–475.

Schraw, G., Dunkle, M.E., Bendixen, L.D., & Roedel, T.D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology, 87*, 433–444.

Sternberg, R.J. (1995). In Search of the Human Mind, Orlando: Harcourt Brace College Publishers.

Sternberg, R.J. (1998). *The triarchic mind: A new theory of human intelligence.* NY: Viking Press.

Tobias, S. & Everson, H.T. (1996). *Assessing metacognitive knowledge monitoring.* College Board Report No. 96-01. College Board, NY.

Tobias, S. (1995). Interest and metacognitive word knowledge. *Journal of Educational Psychology, 87*, 399–405.

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

Tobias, S. (1992). The impact of test anxiety on cognition in school learning. In K. Hagvet (Ed.), *Advances in test anxiety research* (Vol. 7, pp. 18–31). Lisse, Netherlands: Swets & Zeitlinger.

Tobias, S. & Everson, H.T. (2000). Assessing metacognitive knowledge monitoring. In G. Schraw (Ed.), *Issues in the measurement of metacognition* (pp. 147–222). Lincoln, NE: Buros Institute of Mental Measurements and Erlbaum Associates.

Tobias, S. & Everson, H.T. (1998, April) *Research on the assessment of metacognitive knowledge monitoring*. Paper presented at the annual convention of the American Educational Research Association, San Diego.

Tobias, S. & Everson, H.T. (1996, April) *Development and validation of an objectively scored measure of metacognition appropriate for group administration*. Paper presented at the annual convention of the American Educational Research Association, New York.

Tobias, S., Everson, H.T., & Laitusis, V. (1999, April). *Towards a performance based measure of metacognitive knowledge monitoring: Relationships with self-reports and behavior ratings*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Tobias, S., Everson, H.T., Laitusis, V., & Fields, M. (1999, April). *Metacognitive knowledge monitoring: Domain specific or general?* Paper presented at the annual meeting of the Society for the Scientific Study of Reading, Montreal.

Tobias, S., Everson, H.T., & Tobias, L. (1997, March). *Assessing monitoring via the discrepancy between estimated and demonstrable knowledge*. Paper presented at a symposium on *Assessing Metacognitive Knowledge Monitoring* at the annual convention of the American Research Association, Chicago.

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.

Tobias, S. & Fletcher, J.D. (Eds.) (2000). *Training and retraining: A handbook for business, industry, government, and the military*. New York: Macmillan Gale Group.

Touchstone Applied Science Associates (1991). *Degrees of Reading Power*. New York: Touchstone.

Weinstein, C.E. & Mayer, R.E. (1986). The teaching of learning strategies. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed. pp. 315–327). New York: Macmillan.

Weinstein, C.E., Zimmerman, S.A., & Palmer, D.R. (1988). Assessing learning strategies: The design and development of the LASSI. In C.E. Weinstein, E.T. Goetz, & P.A. Alexander (Eds). *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 25–39). New York: Academic Press.

Winne, P.H. (1996). A metacognitive view of individual differences in self regulated learning. *Learning and Individual Differences, 8*, 327–353.
Wright, D. B. (1996). Measuring feeling of knowing. *Applied Cognitive Psychology, 10*, 261–268.
Zimmerman, B.J. & Risenberg, R. (1997). Self regulatory dimensions of academic learning and motivation. In G.D. Phye, *Handbook of academic learning: Construction of knowledge* (pp. 105–125). San Diego: Academic Press.

CollegeBoard
connect to college success™