4. Assessing Metacognitive Knowledge Monitoring

Sigmund Tobias
*Fordham University, sigtobias@gmail.com*

Howard Everson
*The College Board, heverson@collegeboard.org*

Follow this and additional works at: [http://digitalcommons.unl.edu/burosmetacognition](http://digitalcommons.unl.edu/burosmetacognition)

Part of the [Cognition and Perception Commons](http://digitalcommons.unl.edu/cognitionandperception), and the [Cognitive Psychology Commons](http://digitalcommons.unl.edu/cognitivpsychology)


[http://digitalcommons.unl.edu/burosmetacognition/5](http://digitalcommons.unl.edu/burosmetacognition/5)

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Issues in the Measurement of Metacognition by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Assessing Metacognitive Knowledge Monitoring

Sigmund Tobias
Fordham University

Howard Everson
The College Board

Metacognition has been defined as the ability to monitor, evaluate, and make plans for one’s learning (Flavell, 1979; Brown, 1980). Research has shown that learners with effective metacognitive skills are more capable of making accurate estimates of what they know and do not know, of monitoring and evaluating their on-going learning activities, and of developing plans and selecting strategies for learning new material. A large body of literature, reviewed in the other chapters of this volume, has reported differences in metacognitive abilities between learning disabled and regular students, as well as between generally capable learners and their less able counterparts. This research clearly indicates that metacognitive abilities are critically important for effective learning.

Metacognitive processes are usually divided (Pintrich, Wolters, & Baxter, this volume) into three components: knowledge about metacognition, monitoring of metacognitive processes, and control of those processes. The research described in this chapter concentrates...
on the monitoring component of metacognition, specifically students' abilities to monitor their learning by differentiating between the known and unknown. It is assumed that effective control of learning cannot occur in the absence of accurate monitoring. If students cannot distinguish between what they know and do not know, they can hardly be expected to exercise control over their learning activities, or to select appropriate strategies to attain their goals.

INTRODUCTION

Our concern with assessing knowledge and/or ability monitoring is based on reasoning that it is a crucial component of most learning and training contexts. In such situations the learner usually has to master a great deal of new knowledge. Therefore, those who accurately distinguish between what they have already learned and what is yet to be acquired have an important advantage, because they can refrain from studying material that has already been mastered, or merely review it briefly. Such students can then devote most of their time and energies to new, unfamiliar materials. In contrast, those with less effective knowledge monitoring processes are likely to allocate their time and resources less effectively and spend valuable time studying what is known at the expense of the unfamiliar material and, consequently, have greater difficulty mastering new subjects. For these reasons, the program of research described in this chapter concentrated on the development of a procedure to assess students' abilities to monitor their knowledge, and differentiate between what they believe they know and do not know and what they actually know and do not know.

The purposes of this chapter are to describe the metacognitive knowledge monitoring assessment (KMA) we have developed, and to report on a program of research—12 studies in all—that relate scores on the procedure to reading comprehension, problem solving in mathematics, and, more generally, to learning in school settings. Other studies also related scores on the KMA to such variables as anxiety, interest, and need for feedback, and examined the usefulness of the procedure in differentiating among learning disabled, attention deficit hyperactive, and students without special educational needs. All of the studies reported in this chapter used the KMA, a procedure that may be administered using paper and pencil or via computer. The procedure can also be scored objectively and, unlike other assessments of metacognitive processes, it does not rely solely on self-reports of cognitive processing.
Assessing Metacognition

Despite its importance in meaningful human learning, the assessment of metacognition has proven to be both difficult and time-consuming (Pintrich et al., this volume). Metacognition, as a higher order executive process (Borkowski, this volume; 1995), monitors and coordinates the cognitive processes employed during learning. As can be expected, there are considerable difficulties in assessing such higher level processes. Metacognition is usually assessed in two principal ways: observations of students' performance or by self-report inventories. Some problems associated with each of these forms of assessment are described below.

Observation and Verbal Report

Assessing metacognition by observation and verbal reports usually requires all of the following: (a) that students work on some task individually; (b) that their performance is carefully observed; and (c) that their performance is recorded in some way (notes taken by observers or audio/videotapes). Often a number of additional steps are required before a rating of metacognition can be made, including detailed interviews of students, the development of "think aloud" protocols collected as students work on a learning task, and the recording of students' introspective reports. Multiple raters are usually needed to inspect both the records of the performance and the interviews, or introspection protocols, before a sound rating of metacognition can be made (Meichenbaum, Burland, Gruson, & Cameron, 1985). 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). The resources for such work are rarely available in most instructional situations or in many university-based research programs. Pressley's work (this volume; Pressley & Afflerbach, 1995) provides a good example of the complexities of conducting protocol analysis, and Baker and Cerro's chapter (this volume) also discusses some problems with this approach, especially as it pertains to the use of error detection for assessing metacognition.

Labor intensive practices such as those described above make it difficult to evaluate metacognition in many instructionally relevant settings, including secondary and post-secondary schools, as well as training environments in business-industry, governmental agencies, or in the military. In view of these difficulties it is not surprising that most metacognitive research is usually conducted in elementary and some secondary school settings where the time of those participating
in the research can easily be diverted for the research effort. Of course, substantial resources still have to be devoted to enable researchers to collect such metacognitive data.

Self-Report

A number of self-report measures of metacognition (Everson, Hartman, Tobias, & Gourgey, 1991; Jacobs & Paris, 1987; O’Neil, 1991; Pintrich, Smith, Garcia, & McKeachie, 1991; Schraw & Dennison, 1994) have been developed and are widely used. Such questionnaires have the advantage of being easily administered to groups and may be scored rapidly and objectively. Self-report scales usually ask respondents to select from a set of printed choices the cognitive processes and strategies they use while learning from instruction. Such scales put a premium on effective reading abilities and, therefore, are not usually suitable for use with younger or early elementary school children.

Unfortunately, the use of self-report measures in assessing a complex process such as metacognition raises a variety of questions, including some of the following: Because metacognition involves the monitoring, evaluation, and coordination of cognitive processes, are students aware of the processes used during learning? Further, are students able to describe and report on metacognitive processes used, even by merely selecting from available alternatives on a multiple-choice scale? Finally, there is the question of whether students report honestly on the processes. Although the truthfulness of students’ answers is always an issue with self-reports, it may apply especially to reports of cognitive processes used during learning because students at any level are probably reluctant to admit that they may be relatively casual during their attempts to complete school assignments. Of course, these concerns are minimized if appraisals of any construct, and evaluations of metacognition in particular, do not rely on self-reports.

Rationale for Assessing Knowledge Monitoring

Each of the studies reported in this chapter employed a technique for assessing metacognitive monitoring that simultaneously evaluated students’ self-reports of their declarative word knowledge, or their procedural math problem-solving ability, and their demonstrated knowledge or ability. The basic strategy is to assess knowledge monitoring by evaluating the discrepancy between students’ estimates and their actual (determined by performance on a test) knowledge or ability. On the KMA, students are first asked to estimate their knowledge or ability to solve mathematical problems.
The actual knowledge or problem-solving ability is subsequently assessed by administering an objectively scored test, most frequently in multiple-choice format. The discrepancies between students' estimates and their actual knowledge are used as an index of the accuracy of students' metacognitive knowledge monitoring abilities.

The KMA generates four scores that reflect the relationship between students' knowledge estimates and their test performance. Two scores indicate that students estimated knowing an item, or being able to solve a problem, (a) and answered the question correctly on a test (abbreviated as ++), (b) or answered it incorrectly (+ -). Two further scores are generated indicating that students estimated that they do not know an item, or are unable to solve a problem, and (c) answer it correctly (- +), or (d) incorrectly (- -). Of course, the ++ and -- scores are assumed to reflect accurate knowledge monitoring judgments, and the + - and - + scores reflect inaccurate judgments.

Like other types of metacognitive measures, KMA estimates also consist, in part, of self-reports. However, such reports typically are much more readily available to students than the questions usually appearing on self-report inventories dealing with their recollections of the cognitive processes engaged in during learning, and/or how frequently the processes were used. More important, the KMA also incorporates students' actual performances on a test. Because estimated and actual performance can both be scored objectively, the procedure has a clear-cut advantage over asking students to report on their cognitive processes either in the form of protocols, or by choosing from available alternative on self-report inventories.

School assessments are often used to determine whether students learned material presented in class. Therefore, it is important to evaluate students' ability to update their knowledge and make accurate metacognitive estimates of whether the new material was learned, in addition to assessing their prior learning. Consequently, several of the studies reported below also examined students' accuracy in monitoring whether they had mastered materials after being given the opportunity to do so.

The KMA was applied to the domain of students' declarative word knowledge in 10 of the 12 studies described in this chapter. This domain was selected because of its relevance to school learning. In order to demonstrate that the procedure generalizes to other academic domains, two studies dealt with students' procedural knowledge in the area of solving mathematical problems, another important domain in school learning at all levels. Finally, the research described below also examined the relationship of KMA scores and measures of
reading comprehension, school learning, anxiety, interest, and need for feedback, as well as examining whether the KMA differentiated between regular students and those diagnosed as being either learning disabled or having an attention deficit hyperactivity disorder.

Reports of the studies are organized into different categories according to the variables examined. Because a number of the investigations dealt with multiple variables, some studies appear under more than one rubric. In such instances, a detailed report of the study is given when it is first described, and the reader is directed back to that description in subsequent, briefer references to that investigation.

KNOWLEDGE MONITORING AND READING COMPREHENSION

There has been a good deal of research demonstrating that word knowledge or vocabulary is one of the major components of reading comprehension and learning more generally (Breland, Jones, & Jenkins, 1994; Just & Carpenter, 1987). However, few investigations studied whether the accuracy of students’ estimates of their word knowledge was an important predictor of the ability to learn. If students are unable to differentiate accurately between the words they know and do not know, they must find it difficult to determine whether to slow down while reading and try to figure out the meaning of a word from the context, or go to a dictionary to have it defined, or go on in the possibly mistaken or uncertain belief that they understand the word’s meaning. Such uncertainty must be reflected in reduced reading comprehension for students with inaccurate knowledge monitoring. On the other hand, being able to distinguish accurately between words students can define correctly and those they cannot should enhance their reading comprehension and their effectiveness in learning new material. Because a great deal of research on metacognition has dealt with reading comprehension, the criterion for assessing the validity of the KMA in the first two studies was to determine its relationship to measures of reading comprehension.

Study I: Estimates of Word Knowledge and Reading Comprehension

In view of the demonstrated relationships between metacognition and reading comprehension, it seemed important to evaluate the accuracy of students’ monitoring of their word knowledge in a

reading context. Such a setting was expected to increase the relevance of the assessment to school learning. It was also anticipated that the ability to learn new vocabulary would be an important skill for reading specifically, and school learning more generally. Furthermore, students' abilities to make accurate metacognitive assessments of whether they had actually learned the meanings of new words, given an opportunity to do so, would also seem to be an important indicator of reading comprehension. Therefore, the ability to update one's knowledge and to make metacognitive estimates of the updated knowledge were also assessed in this study.

Participants and Procedures

Participants were randomly assigned to one of two conditions. One group was asked to read a 750-word text passage, and then complete a word list and vocabulary test composed of words that had been defined explicitly or implicitly in the text. The second group received the Sentence Verification Test (SVT; Royer, Lynch, Hambleton, & Bulgarelli, 1984) rather than the text passage as a control. The text passage described the incidence and prevalence of heart disease, the risk factors for developing heart ailments, the technical terms for varying degrees of the illness, the characteristics differentiating the different degrees, and a number of ways by which the risks of developing heart disease could be reduced. It was known from prior research (Tobias, 1989; 1969) that there was a good deal of variability in participants' prior knowledge of this material.

On the word list, participants were asked to indicate, by checking off one of two blanks, whether they knew, or did not know each of 33 words. All of the words were defined, either explicitly or implicitly, in the passage previously administered to the group who read the heart disease text. When the word list was completed, students received a four-choice vocabulary test containing all of the 33 items on the word list with instructions to select the correct synonyms or definitions of the words. A number of other research instruments were also administered, as was the Descriptive Test of Language Skills, Reading, and Comprehension (DTLS; College Board, 1979), a standardized test of reading comprehension.

The text passage, word list, and vocabulary test were examined by four raters who judged whether the words were defined implicitly or explicitly in the text. The passage was revised until consensus was reached among the judges. Of the 33 words, the ratings indicated that 25 were defined implicitly (e.g., "Epidemiologists who have com-
pared the prevalence of heart disease in the United States and in other countries...”) and eight words were defined explicitly (e.g., “Coronary or heart disease....”).

A total of 167 freshmen at a large urban university participated in this study. The students attended a summer session program designed to familiarize them with the university and the skills needed to succeed in their studies. The group receiving the SVT consisted of 87 students, and 82 subjects read the text passage.

Results and Discussion

The accuracy of students’ metacognitive word knowledge judgments was determined by comparing students’ estimates of their knowledge with their performance on the vocabulary test. The four scores described earlier were generated: Terms checked as being known on the word list which were scored (a) correct [abbreviated as + +], or (b) wrong [shown as + - ] on the vocabulary test. Two further scores described terms students checked as being unknown on the word list and answered (c) correctly [abbreviated as - +], or (d) incorrectly [- -] on the vocabulary test.

The four KMA scores were computed for the total set of words, and also for those that were defined explicitly or implicitly. The correlations between these data and the reading comprehension subtest of the DTLS are shown in Table 1, for all subjects combined, as well as for the group receiving the Sentence Verification Procedure (SVT) and those reading the heart disease text.

The correlations in Table 1 indicate that, as expected, accurate metacognitive judgments about the number of words students thought they knew and answered correctly on the test (T+ +) had a substantial positive relationship with reading comprehension. Estimates of the number of words thought to be unknown and answered incorrectly (T- -) were negatively related to comprehension. Furthermore, and also anticipated, accurate estimates of words defined explicitly (E + + and E - -) and implicitly (I + + and I - -) were also significantly correlated (see Table 1) with comprehension, whereas the incorrect judgments (E + -, E - +, I + -, and I - +) were not. The magnitude of many of the correlation coefficients is especially impressive because the participants were relatively homogeneous with respect to ability, because they were considered to be at risk of doing poorly in school and, therefore, advised to participate in the orientation and pre-freshmen skills program they were attending.

The relationships between the KMA scores and reading comprehension were dramatically lower for students who did not read the
text passage, and received the SVT, compared to the others. Those reading the passage had the chance to learn the meanings of previously unknown words, or to update their knowledge of familiar and partially known words, whereas students who received the SVT did not have that opportunity. It was expected that students who could update their knowledge would make more accurate metacognitive judgments than the others. Operationally then, it was expected that group membership (i.e., reading the heart disease passage or the SVT) and accuracy of metacognitive judgments would have an interactive effect on reading comprehension. This hypothesis was tested by multiple regression analysis in which a binary vector for group membership (those reading the heart disease passage or the SVT), KMA score, and their product (representing the interaction term), were entered as independent variables and the reading comprehension test score was the dependent variable. The results of that analysis are shown in Table 2.

As expected, the \( t \) test on the beta weights for the interaction term was significant in five of the six equations shown in Table 2. These
Table 2. Beta Weights and Associated $t$ Tests for all Effects on all Derived Scores.

<table>
<thead>
<tr>
<th>Score</th>
<th>Group</th>
<th>Beta</th>
<th>$t$</th>
<th>Beta</th>
<th>$t$</th>
<th>Group X Score</th>
<th>Beta</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T++</td>
<td></td>
<td>-.10</td>
<td>.44</td>
<td>-.62</td>
<td>3.33**</td>
<td>.88</td>
<td>2.67**</td>
<td></td>
</tr>
<tr>
<td>T-</td>
<td></td>
<td>.12</td>
<td>.97</td>
<td>.04</td>
<td>.19</td>
<td>-.54</td>
<td>2.32*</td>
<td></td>
</tr>
<tr>
<td>E++</td>
<td></td>
<td>-.42</td>
<td>3.00**</td>
<td>-.34</td>
<td>1.22</td>
<td>.85</td>
<td>2.60*</td>
<td></td>
</tr>
<tr>
<td>E-</td>
<td></td>
<td>.07</td>
<td>.56</td>
<td>.03</td>
<td>.14</td>
<td>-.41</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>I++</td>
<td></td>
<td>-.59</td>
<td>3.14**</td>
<td>-.07</td>
<td>.29</td>
<td>.82</td>
<td>2.55*</td>
<td></td>
</tr>
<tr>
<td>I-</td>
<td></td>
<td>.10</td>
<td>.84</td>
<td>.00</td>
<td>.02</td>
<td>-.50</td>
<td>2.17*</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05  
**p<.01  
T = Results for total word list.  
E = Results for words defined explicitly.  
I = Results for words defined implicitly.  
++ = Words students claimed to know and got right on vocabulary test.  
+ = Words students claimed to know and got wrong on vocabulary test.  
- = Words students claimed not to know and got wrong on vocabulary test.  
+ = Words students claimed not to know and got right on vocabulary test.

results indicate that students who could update their word knowledge by reading the passage made significantly more accurate metacognitive judgments than those who did not have that chance. This finding is not surprising because the major skill assessed for the group reading the passage was probably the ability to infer the meaning of words, surely an important component of reading comprehension. Clearly then, the opportunity to renew word knowledge and then estimate mastery of the updated knowledge improved the relationships with reading comprehension.

Estimates and Number Correct. The metacognitive scores described above were a function of two factors: Knowledge as reflected in the number of items students answered correctly on the vocabulary test, and knowledge estimates seen from how accurately students estimated that number. One question that arises is whether students' knowledge estimates contributed variance above and beyond their actual knowledge reflected by the total number correct, or raw score. Of course, a great deal of research has demonstrated that students' vocabulary scores are highly related to reading comprehension and
school learning more generally (Breland, Jones, & Jenkins, 1994; Just & Carpenter, 1987). In the KMA, the raw score may be obtained by adding the ++ and -+ scores. For the monitoring procedure to be useful, it should account for more variance than the number of items students answered correctly, irrespective of their knowledge estimates. That question is examined below in the first study and for all of the other investigations described in this chapter.

The correlation between the raw score on the vocabulary test (total number of words correct) and the DTLS was .45. As Table 1 indicates, the highest relationship among the metacognitive estimates and reading ability, $r = .65$, was between the total number of words estimated to be known and actually known ($T++$). The difference in the magnitude of these correlations indicates that accurate estimates of students' word knowledge contributed variance above and beyond the total vocabulary score. When $T++$ was forced into a regression equation, the total number of words correct, irrespective of prior estimates, did not contribute enough independent variance to enter the equation. That finding confirms the differences between the two correlation coefficients described above, and indicates that the accuracy of students' estimates of their updated vocabulary knowledge were more highly related to comprehension than the actual knowledge.

The results of this first study were encouraging with respect to the construct validity of the KMA. As expected, metacognitive assessments of students' word knowledge were more substantially related to reading comprehension than the number of correct answers alone.

Study II. Declarative Word KMA and Reading Comprehension

The preceding study found strong relationships between metacognitive monitoring and reading comprehension in general. The purpose of the second study was to determine the KMA's relationship both to prior reading ability and some of the components of reading comprehension, such as identifying words in context, understanding meaning, and understanding the writer's tone and assumptions. The four basic KMA scores appeared to have some similarity to the phenomena studied in signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991), which separates phenomena into signal and noise components. Therefore, a further purpose of this study was to examine whether the signal

---

3This study, by Howard Everson, Ivan Smodlaka, and Sigmund Tobias was published in *Stress, Anxiety, and Coping*, 1994. See References.
detection paradigm could define more useful scores than the ones used in Study I. Finally, relationships between KMA scores and measures of test anxiety were examined, and are reported later in this chapter.

Participants and Procedures

The word list and vocabulary test used in the first study were administered to students, together with the Worry subscale of the Test Anxiety Inventory (Spielberger et al., 1980), and the Descriptive Test of Language Skills, Reading, and Comprehension (College Board, 1989) that contained three subscales: identifying words in context, understanding meaning, and understanding the writer’s tone and assumptions. An archival index of reading ability was obtained from the participants’ school records. The participants were 117 undergraduates from a large urban university, 65% were women.

Results and Discussion

Knowledge monitoring ability was assessed by computing “hits” (i.e., the number of words each participant claimed to know and subsequently identified correctly on the vocabulary task or conversely said they did not know and failed to identify correctly on the subsequent vocabulary task), and “false alarms” (i.e., the number of words each claimed to know but did not correctly identify, and those claimed to be unknown yet correctly identified). Using signal detection theory, these two indices were transformed into a d’ index that provides an estimate of metacognitive sensitivity and B, an index that provides an estimate of the participants’ response bias. The reliability estimate (Cronbach, 1951) of these two indices was .78.

In general, the more capable readers demonstrated higher levels of metacognitive ability. The correlations of knowledge monitoring ability—as measured by the d’ index—with prior reading ability and the experimental measure of reading comprehension were .35 and .39, respectively. Moreover, hierarchical multiple regression analyses permitted us to isolate the effects of metacognitive ability on reading test performance, once prior reading ability and anxious worry were controlled statistically. These analyses suggested that metacognitive ability was positively related to reading test performance (B = .17, t = -2.23, p = .03). Similarly, the correlations with the reading test’s subscales measuring vocabulary in context, literal interpretation of text, and understanding the writers’ tone and assumptions were .32, .43, and .26 respectively.
Contrasts with Study I

In Study II the text in which all of the vocabulary words were defined was not administered. The correlation of .35 between the d’ score and reading comprehension was similar to the correlation of .29 (see Table 1) found in the first study between T++ and reading comprehension for those students who did not read the text passage. Of course that relationship is much lower than the correlation of .65 found in Study I between the same variables for students reading the passage. Clearly then, these two studies suggest that the metacognitive word knowledge scores derived from the KMA had a strong, consistent relationship with standardized measures of reading comprehension and, further, that the opportunity to renew word knowledge and re-estimate mastery of the updated knowledge improved the relationships with reading comprehension.

KNOWLEDGE MONITORING AND SCHOOL LEARNING

The first two studies were encouraging with respect to the relationship of the declarative word KMA to reading comprehension. The results of these investigations indicated that metacognitive estimates were closely related to competence in the domain in which students’ estimates of knowledge were obtained (i.e., reading). One purpose of the studies described below was to examine whether the declarative word KMA was related to a more distant domain than the one in which the assessment occurred, such as learning in school. The expected relationship with school learning seemed reasonable because accurate estimates of one’s knowledge should make it easier to acquire the large amounts of new information taught in such settings. Four studies dealing with these questions are described below. Furthermore, because the vocabulary and text passage dealt largely with familiar issues and had a minimal technical vocabulary, the task of inferring the meanings of unknown words from the passage, or estimating one’s word knowledge seemed most similar to learning in courses that rely largely on conventional vocabulary, rather than introducing a large set of new technical terms. Therefore, it seemed likely that declarative word KMA scores should be more closely related to students’ learning in English and Humanities courses than in others.

Another purpose of the succeeding studies was to extend the research on metacognitive knowledge monitoring to the learning of students in secondary and post-secondary schools. As mentioned above, much of the research dealing with metacognition has been
conducted in elementary schools, and to a lesser degree in secondary or post-secondary school settings. Two of the succeeding studies examined the relationship of the knowledge monitoring procedure to students' overall achievement in college, and to their learning in different content domains, and two others used high school students and those who dropped out of school.

Study III. Knowledge Monitoring and College Learning

Students acquire a great deal of new knowledge in secondary and post-secondary schools. Therefore, their ability to estimate whether they have mastered either previously learned content or new material seemed to be an important characteristic of effective learners, especially in college. Accurate monitoring of new learning should enable students with effective knowledge monitoring strategies to concentrate on new materials and skim over familiar content. On the other hand, students with less effective knowledge monitoring may waste time practicing or reviewing what they already know, rather than zeroing in on new material or updating partially learned content. Therefore, Studies III and IV asked students to estimate their vocabulary knowledge twice: the first time to assess their prior learning, and the second to determine their ability to update prior learning. It was assumed that students' accuracy in estimating their word knowledge after having the chance to update it would be more closely related to college learning reflected in their grade point average (GPA) than to estimates of prior learning.

The word list, vocabulary test, and text materials used in the two studies reported above contained a much larger set of explicitly defined words compared to those defined implicitly. It was reasoned that implicit definitions might be especially important for college learning, where students frequently had to infer the meanings of new words from context. Therefore, the materials were modified to increase the number of implicitly defined words.

Participants and Procedures

The word list, vocabulary test, and text passage were revised to contain an equal number of target words that were defined explicitly and implicitly in the text passage. The expository text used in one of the prior studies was revised and a narrative version of the same passage was developed in order to examine the effect of situational

---

4A paper based on Studies III and IV was presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April 1995.
interest on metacognitive knowledge monitoring (findings dealing with interest are discussed later in the chapter).

The word list and vocabulary test contained 38 words, half were explicitly defined and the others received implicit definitions. Types of definitions were determined by two independent judges who rated all words. Disagreements were resolved by revising the passage until agreement was reached. Because these materials are used in six of the studies described in this chapter, a sample, consisting of the first page of the materials, is shown in Figures 1-3.

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

Figure 1. Word List for Knowledge Monitoring Procedure.

Please indicate whether you know, or do not know each of the words listed below, by checking the appropriate space.

<table>
<thead>
<tr>
<th>Word</th>
<th>Know</th>
<th>Do not know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abuse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascribed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attenuate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benign</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholesterol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deterrent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emanating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epidemiology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Esoteric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etiology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatalities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gravity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guarded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implicated</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For each word check the space which means most nearly the same thing as the first word.

1) Prevalent
   - a) stronger
   - b) winning
   - c) frequent
   - d) prior

2) Attributed
   - a) caused
   - b) ovation
   - c) stream
   - d) tax

3) Optimal
   - a) best
   - b) opening
   - c) eyelid
   - d) cheerful

4) Obesity
   - a) listen
   - b) fat
   - c) apology
   - d) obsolete

5) Acute
   - a) pretty
   - b) serious
   - c) heavy
   - d) often

6) Ascribe
   - a) refer
   - b) written
   - c) question
   - d) bed

7) Transitory
   - a) move
   - b) temporary
   - c) carry
   - d) train

8) Median
   - a) stripe
   - b) divider
   - c) middlemost
   - d) negotiate

9) Ingest
   - a) joke
   - b) eat
   - c) enter
   - d) exit

10) Residual
   - a) lasting
    - b) live
    - c) income
    - d) clever

11) Infarction
   - a) tooth decay
    - b) particle
    - c) rule violation
    - d) muscle death

12) Fatalities
    - a) fatty tissue
    - b) deaths
    - c) fateful
    - d) take in stride

13) Incidence
    - a) new cases
    - b) an example
    - c) exciting
    - d) event

14) Attenuate
    - a) listen
    - b) reduce
    - c) pay attention
    - d) try

15) Guarded
    - a) uncertain
    - b) optimistic
    - c) degrees
    - d) watchful

The sample consisted of 139 students attending a large urban university, though only 84 subjects completed all the materials during two sessions. Part of the sample consisted of students entering the nursing program (N = 47, N = 33 with complete data) who were taking an orientation course in that department. The rest of the sample consisted of freshmen (N = 92, N = 51 with complete data) taking a freshman orientation course.

Results and Discussion

The correlation between total score on both administrations of the vocabulary test, based on 84 students who completed the test on both administrations, was .75. This is not a test-retest reliability coefficient.
Coronary or heart disease is a major health problem among all ethnic, racial and occupational groups in the United States. In addition to coronary disease, health workers are worried about many other maladies affecting Americans, such as cancer, AIDS, and other equally serious conditions. However, compared to all other serious illnesses, coronary problems cause more than half of the total number of fatalities or deaths in the United States. To be exact, 55% of the deaths among all groups in this country, or more fatalities than for all the other illnesses combined, may be ascribed to coronary disease. Not only is coronary disease responsible for the greatest number of fatalities in this country but it is also the most prevalent, or frequent, of all the serious illnesses. That is, coronary disease is more prevalent than all the other serious conditions combined.

The incidence, that is the number of new cases, of coronary disorders is higher among men than among women for the country as a whole. The incidence of heart disorders is also higher for cigarette smokers than it is among non-smokers. A higher incidence of coronary disease among Americans is also attributed to alcoholism, drug addiction, and tobacco. The etiology, or causes, of coronary disease among Americans are not completely clear, but excessive use, or abuse of alcohol and the other substances mentioned above is often linked to coronary disease. In addition, tension, air pollution, weighing too much, and engaging in too little exercise are also implicated as causes of heart disease among people living in the United States.

The gravity of heart disease for people in general is a function of the magnitude of coronary damage. The heart is basically a muscle similar to all the others in the human body. The amount of damage to the heart muscle, or myocardium, determines the seriousness of the illness. The most serious type of damage, which is called myocardial infarction, occurs when the heart muscle dies. One major difference between the myocardium and other muscles in the human body...
because students read the text passage, from which the meaning of the words could be inferred, immediately before the second administra-

Students’ estimated word knowledge and performance on the vocabulary test were determined for both administrations. Two scores were computed for each administration: the total number of correct [words in the ++ and -- categories] and incorrect [+ - and - + categories] estimates. Preliminary analysis found no differences between students assigned to the expository or narrative text versions, or between explicitly and implicitly defined words, therefore the data for both text versions and both types of words were pooled. The correlations between the correct and incorrect estimates on both administrations of the words and students’ GPA in English, Humanities, Sciences, Social Sciences, and combined GPA were computed and are shown in Table 3. Because 92 participants were freshmen in their first term of college the overall GPA for this group was based on an average of only 12.1 credits (SD = 5.6), whereas the nursing students had a mean of 56.4 credits (SD = 28.3). Therefore, the correlations are presented for each group separately, as well as for the total sample. Table 3 also shows the correlations for metacognitive estimates and raw score, number correct on the vocabulary, separately. Finally, the different number of cases in the various cells of Table 3 should also be noted.

The correlations shown in Table 3 are generally positive and frequently significant, even though they ranged from low to moderate in magnitude. The results support the concurrent validity of the procedure with respect to its relationship to learning in college. As expected, correlations between knowledge monitoring scores and GPA in English were generally highest; presumably accurately estimating word knowledge is more important in English than in other subjects. Relationships with Humanities courses and with the combined GPA were generally significant and lower than those with English grades; correlations with social science and science GPAs were generally lower, and usually not significant. The largely nonsig-
ificant relationships with social and behavioral science courses were surprising because it had been assumed that these courses usually contained less technical or unfamiliar material and vocabulary than the sciences. Perhaps grades in these courses, like those in science, reflected greater domain specific knowledge than found in English and Humanities classes.

The significance levels of the correlations reported in Table 3 varies widely, probably as a function of at least three factors. First, the number of cases in each cell differs due to students’ absence from
Table 3. Correlations Between Knowledge Monitoring Procedure Scores, Raw Scores, and Overall Grade Point Averages in Different Subject Areas.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Administration 1</th>
<th>Administration 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct Estimate</td>
<td>Raw Scr</td>
<td>Correct Estimate</td>
<td>Raw Scr</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total GPA</td>
<td>n</td>
<td>r</td>
<td>n</td>
<td>r</td>
</tr>
<tr>
<td>Total</td>
<td>101</td>
<td>.20*</td>
<td>94</td>
<td>.09</td>
</tr>
<tr>
<td>Freshmen</td>
<td>65</td>
<td>.09</td>
<td>61</td>
<td>-.10</td>
</tr>
<tr>
<td>Nurses</td>
<td>36</td>
<td>.28*</td>
<td>33</td>
<td>.19</td>
</tr>
<tr>
<td>English GPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>72</td>
<td>.30**</td>
<td>63</td>
<td>.19</td>
</tr>
<tr>
<td>Freshmen</td>
<td>53</td>
<td>.31**</td>
<td>48</td>
<td>.00</td>
</tr>
<tr>
<td>Nurses</td>
<td>19</td>
<td>.25</td>
<td>19</td>
<td>.45*</td>
</tr>
<tr>
<td>Humanities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>82</td>
<td>.26**</td>
<td>74</td>
<td>.13</td>
</tr>
<tr>
<td>Freshmen</td>
<td>52</td>
<td>.12</td>
<td>46</td>
<td>-.11</td>
</tr>
<tr>
<td>Nurses</td>
<td>30</td>
<td>.48**</td>
<td>28</td>
<td>.35*</td>
</tr>
<tr>
<td>Science GPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>65</td>
<td>.18</td>
<td>63</td>
<td>.03</td>
</tr>
<tr>
<td>Freshmen</td>
<td>28</td>
<td>.11</td>
<td>27</td>
<td>-.28</td>
</tr>
<tr>
<td>Nurses</td>
<td>37</td>
<td>.26</td>
<td>36</td>
<td>.18</td>
</tr>
<tr>
<td>Social Science GPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>.18</td>
<td>63</td>
<td>.24*</td>
</tr>
<tr>
<td>Freshmen</td>
<td>26</td>
<td>.15</td>
<td>29</td>
<td>.14</td>
</tr>
<tr>
<td>Nurses</td>
<td>38</td>
<td>.09</td>
<td>34</td>
<td>.14</td>
</tr>
</tbody>
</table>

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

either the first or second administration of the materials. Second, it is well known that college grades are often unreliable (Werts, Linn, & Joreskog, 1978; Willingham, Lewis, Morgan, & Ramist, 1990), reducing the magnitude of any correlations with them. Third, students completed a varying number of courses in each area, thus GPAs may have been based on one or a few courses in some fields, reducing the stability of the criterion. The reliability of the grades may have been reduced further by three factors: (a) students took dissimilar courses in each of the areas shown in Table 3; (b) when similar courses were
taken they were taught by different instructors; and (c) differences in students' major fields of study.

As expected, the correlations between knowledge monitoring scores and grades in English were generally higher, and more frequently significant, than those of any other subject. For the 84 students with complete data for both administrations of the vocabulary test, the mean total score increased from 23.3 ($SD = 6.0$) for the first vocabulary test to 26.0 ($SD = 6.6$) for the second ($t(83)= 5.53, p < .001$). Thus students clearly learned the meanings of some of the words after having the chance to update their word knowledge by reading the passage. However, in contrast to the results of the first study, the relationships between the metacognitive scores and grades shown in Table 3 were generally higher before students read the text passage than afterwards. The Study I findings of higher relationships with DTLS scores on the second administration of the procedure may be attributable to the use of reading comprehension scores rather than grades as criteria. Apparently, inferring the meaning of words is a more important component of reading comprehension than of school learning more generally.

It was assumed in this study that having the chance to update one's word knowledge before estimating it would be more similar to students' learning in their classes than merely estimating prior word knowledge. Therefore, relationships with grades were expected to be higher for the second administration than the first. The findings did not support these expectations. Although the increase in vocabulary score after reading the text was statistically significant, it indicated that, on average, less than three new words were learned from the text passage. Perhaps such modest acquisition was dissimilar to the amount of learning in college courses leading to lower relationships with metacognitive monitoring scores on the second administration of the procedure. Similarities between the knowledge monitoring task and school learning might have increased if students were instructed to study the passage more intensely, or asked to pay special attention while reading words they had previously seen on the vocabulary test. Such instructions may have increased the correlations with GPA for the second administration. It remains for further research to explore that possibility.

Table 3 also indicated that the correlations with number correct on the vocabulary test were generally similar to the relationships with correct knowledge monitoring estimates. Due to the varying $Ns$ in the different cells, the significance of differences in correlations was examined with a $t$ test developed by Hotelling (1931). For the
correlations with GPA based on both administrations, using the total group, the knowledge monitoring scores were higher seven times (one difference was significant at \( p < .05 \)), and the correlations based on raw scores were higher three times (none significantly so). For freshmen, the correlations with knowledge monitoring procedure scores were higher twice, but not significantly so, and correlations with raw score were higher eight times (two significant at \( p < .05 \)). Finally, for nursing students, correlations based on knowledge monitoring scores were higher five times (none significant), and relationships based on raw scores were higher five times (one significant \( p < .05 \)). Thus, the knowledge monitoring scores appeared to add little independent variance to the relationship with grades beyond that accounted for by the number correct on the vocabulary test.

The findings for this study, in contrast to the findings of the first two investigations, suggest that the knowledge estimating procedure seems to account for little independent variance in GPA above that attributable to number correct on the vocabulary test. Conceivably the findings of low reliability for college grades (Werts, Linn, & Joreskog, 1978; Willingham, Lewis, Morgan, & Ramist, 1990), referred to above, may have contributed to these findings. The criterion in the first two studies consisted of test scores, which are much more reliable than grades.

Study IV: Predicting College Learning from KMA Scores

The preceding study dealt with the concurrent validity of the KMA by examining the correlations of knowledge monitoring procedure scores with students’ prior learning in college. The fourth study investigated the KMA scores’ predictive validity by examining whether metacognitive estimates predicted entering students’ performance during their first year of college.

Participants and Procedures

The materials used were identical to those described in Study III. They were administered while students attended a prefreshman skills program prior to beginning their first semester of college. Learning was determined by obtaining students’ GPAs at the end of their first college year in the same areas examined in the prior study: English, Humanities, Science, and Social and Behavioral Science, as well as the combined GPA. The sample consisted of 115 students (59 female) participating in a prefreshman skills program intended for students considered at risk of doing poorly in their first year of college.
Results and Discussion

The number of correct metacognitive estimates of students' word knowledge were determined. As in the prior studies, correct estimates were defined by combining the ++ and -- categories. Preliminary analysis again indicated that there were no differences between the expository and narrative passages, nor between the words defined explicitly or implicitly. Therefore, these data were pooled for the succeeding analyses.

Correlation analysis was the optimal data analytic mode in the preceding study because of the large amount of missing data due to student's absences, and the varying courses in different areas taken by the freshmen and prenursing students. However, by examining whether changes in knowledge monitoring scores were accompanied by similar changes in GPA, correlations were likely to maximize errors attributable to the low reliability of grades because small changes that could alter the correlations might be attributable to error. The participants in Study IV were incoming freshmen who completed all the materials and took similar types of courses. Therefore, high and low achievement groups were created by splitting students at the GPA median in the different academic areas, and on the combined GPA, and then examining knowledge monitoring differences between the groups. Mixed between and within subjects analyses of variance were then computed to determine the significance of differences between the first and second administrations, and of differences in metacognitive estimates between groups above and below the GPA median.

A search of the college records found that 95 of the 115 students examined a year earlier had completed some courses at the school. ANOVA indicated that, as expected, students above the median GPA ($N = 48$) made significantly more accurate overall metacognitive judgments ($\text{Mean} = 49.2$, $F(1, 93) = 6.42$, $p < .05$) on both administrations than those below the median ($N = 47$, $\text{Mean} = 45.8$); the size of that effect, determined by $\eta^2$ (SPSS, 1993), was .065. Also as expected, there was a significant difference between the first ($\text{Mean} = 22.9$) and second administration ($\text{Mean} = 24.5$) of the word list and vocabulary test ($F(1, 93) = 14.95$, $p < .01$, $\eta^2 = .138$), though there was no interaction between these variables. A similar analysis was computed using the number right on both administrations of the vocabulary test as the dependent variable. That analysis indicated that the differences between the high ($\text{Mean} = 43.2$) and low GPA group ($\text{Mean} = 39.3$) on the vocabulary test was not significant ($F(1, 93) = 1.10$, $p = .29$).
2.73, \( \eta^2 = .029 \)), and the differences between the first (Mean = 17.7) and second administrations (Mean = 24.5) were significant (\( F(1,93) = 198.04, p < .001, \eta^2 = .68 \)); again there was no interaction.

High and low groups in English, Humanities, Science, and Social Science courses were also formed by splitting the students at the GPA median in each of these content areas and examining the significance of differences on the number of correct metacognitive estimates. In English, the overall differences in metacognitive accuracy between students above (Mean = 48.9) and below the median (Mean = 45.4) were significant (\( F(1,82) = 6.18, p < .02, \eta^2 = .07 \)), as were the differences between the first (Mean = 45.6) and second administrations (Mean = 48.7, \( F(1,82) = 11.92, p < .01; \eta^2 = .127 \)). Furthermore, there was an interaction between groups and administrations (\( F(1,82) = 4.41, p < .05; \eta^2 = .051 \)). The interaction, shown in Figure 4, suggests that although both groups increased their accuracy from the first to the second vocabulary test in estimating which words were known and unknown, higher achieving students had greater gains than the others. A similar analysis was

![Figure 4. Interaction of GPA Groups, Hits, and Administration.](image-url)
computed for number correct on both vocabulary test administrations. The finding indicated a slightly smaller difference between the high (Mean = 42.9) and low GPA group (Mean = 38.9, \( F(1,82) = 5.43; \eta^2 = .062 \)) than obtained when the metacognitive scores were used, but a stronger effect for differences between first (Mean = 18.0) and second administrations (Mean = 23.6, \( F(1,82) = 169, p < .001; \eta^2 = .673 \)); there was no evidence for interaction in these results.

Similar analyses were computed for students above and below the median in Humanities courses (Art, History, Music, Philosophy, World Civilization, World Humanities, and World Arts). Differences between High (Mean = 49.4) and Low Humanities GPA groups (Mean=45.3) were also significant (\( F(1, 81) = 7.96, p < .01; \eta^2 = .089 \)), as were the differences between first (Mean = 23.0) and second administrations (Mean = 24.5, \( F(1, 81) = 9.94, p < .001; \eta^2 = .109 \)), there was no interaction. The same type of analysis was also computed for number correct on the first and second vocabulary test again revealing somewhat smaller differences between the high (Mean=43.1) and low groups (Mean = 39.0, \( F(1, 81) = 4.18, p < .05; \eta^2 = .049 \)) and larger differences between the first (Mean = 17.8) and second administration (Mean = 23.4, \( F(1, 81) = 179.2, p < .001; \eta^2 = .689 \)) than the results for knowledge monitoring scores. There were no significant differences between the Science or Social and Behavioral Science GPA groups using either the knowledge monitoring procedure or raw scores.

The relationships between metacognitive scores and GPA a year later were generally similar to those reported in Study III, supporting the predictive validity of the KMA scores. Unlike the prior study, in which both knowledge monitoring and raw scores had fairly similar patterns of relationships, the metacognitive scores had a significant effect on overall GPA, whereas the raw scores did not. Furthermore, the knowledge monitoring scores accounted for more variance than the number right in two of three other comparisons, supporting the construct validity of the procedure.

Several factors are likely to have reduced the magnitude of the effects and the generalizability of the results to other college groups. As was the case in the first study, the participants in the pre freshmen program were considered to be at risk for poor performance in college. This factor may have reduced the range of college achievement for the sample and, therefore, may also have reduced knowledge monitoring differences between the groups. Furthermore, even though data were not collected in sections of the pre freshmen skills program devoted exclusively to English as a Second Language (ESL),
some of the students were signed up for both ESL and other skills sections, and thus ended up as part of the sample. The presence of nonnative English speakers could also have reduced the variability among participants and narrowed group differences in this study. Further research limited to native English speakers, who are more heterogeneous in academic skills than the present sample, is needed to determine whether metacognitive differences between low and high achieving students are greater than those reported here.

In general, KMA scores seemed to differentiate the more capable students, whose grades were above the median, from those less able more successfully than did the raw scores, replicating the findings of Studies I and III. The knowledge monitoring scores accounted for anywhere from 1% to 4% more variance than similar analyses using the raw score. It was also interesting to note that the analysis of raw score differences between the first and second vocabulary test administrations always accounted for substantially more variance than did a similar analysis based on knowledge monitoring scores. The latter finding is reasonable and supports the construct validity of the procedure because most students learned some new words from the text passage, though their knowledge monitoring was not equally enhanced. However, it should be noted that the results for English grades indicated that there were greater increases in knowledge monitoring accuracy for capable students than for their less able peers (see Figure 4). These findings suggest that although all students increased both their knowledge and knowledge monitoring accuracy from first to second administration, the increases in metacognitive knowledge monitoring accuracy were greater for more capable students (i.e., those whose English grades were above the median). Apparently such students’ metacognitive skills improved to a greater degree than those of their less able colleagues.

It should be noted that many of the students in this sample took less than a full-time schedule of courses. That is likely to have decreased the reliability of the GPA because it was based on fewer courses and credits than is usually the case after a year of college. This factor may also limit the generalizability of the results to other groups, in addition to reducing the magnitude of the findings by decreasing the potential variability of the GPA. Therefore, in order to increase both the reliability and variability of the criterion, it would be useful to investigate the predictive validity of the knowledge monitoring procedure in settings with a greater percentage of full-time students.
Study V: Knowledge Monitoring and Learning Among Vocational High School Students\textsuperscript{5}

All of the prior studies used college students as subjects. College students are probably more academically oriented than those attending vocational high schools, and consequently more likely to be reflective about their declarative word knowledge and in turn, likely to make more accurate estimates of what they know and do not know. Therefore, one purpose of the next study was to examine the relationship of the knowledge monitoring procedure for students attending a vocational high school. This study also examined the relationship between metacognitive monitoring scores and students' estimates of their performance, as well as their test anxiety. The results dealing with those variables are discussed in later sections of this chapter.

Participants and Procedures

This study employed the word list and vocabulary test described in the two preceding studies; the text passage was not used. Students were tested in one of their regular school classes. In addition some anxiety scales were administered and students were also asked to estimate their grades on tests given in one of their vocational classes. Students' overall GPAs were obtained from the school's permanent records.

All of the participants attended a vocational high school in a large urban school system. A total of 61 students (59 male) participated in this study. The students' ages ranged from 16-19.

Results and Discussion

Students were split at the GPA median, and two multivariate analyses of variance (MANOVA) computed. The first examined differences between the high and low GPA groups on the accuracy of students' knowledge estimates (using the + +, + -, - +, and - - scores) and the second analysis examined group differences in student word knowledge (the sum of ++ and + + scores equal the number correct on the vocabulary test). MANOVA indicated that the overall knowledge monitoring differences between the high and low GPA groups were significant (Transformed Wilks lambda [3,57] = 3.17, \( p < .05 \), effect size = .143). Univariate analyses found that only the difference between the high (Mean = 17.8) and low (Mean = 14.4) GPA groups on the ++ scores were significant ([1,59] = 9.35, \( p < .01 \)).

\textsuperscript{5} The data for this study were collected by Deno Charalambous.
The MANOVA computed on group differences in the number correct on the vocabulary also indicated a significant difference between the groups (Transformed Wilks lambda $F(2,58) = 5.35, p < .01$, effect size = .156). Univariate analyses found that the differences in + + scores were the same as in the preceding analysis; however, in this analysis group differences between the high (Mean = 3.5) and low (Mean = 5.2) GPA groups on the - + scores were also significant ($F(2,58) = 5.59, p < .05$). As expected, the results indicated that the more capable students estimated and actually knew more words than those with lower GPAs; the latter group estimated not knowing more words than the students who were above the GPA median.

The significant differences between the two GPA groups replicate the results of the two prior studies dealing with college GPAs and confirms the relationships between metacognitive knowledge estimates and school learning. The results of the second analysis fail to support the additional importance of obtaining students' estimates of their knowledge because the differences between the GPA groups on their actual vocabulary knowledge were also significant and slightly greater than the differences in knowledge monitoring. The word list and vocabulary test were developed for a college population; perhaps these materials were so unfamiliar to these vocational high school students that their estimates were based on little more than chance.

Study VI. Knowledge Monitoring Among High School Dropouts

The high percentage of students who drop out of high school before graduating is a major problem, especially in times when entry level employees for most positions in business and industry call for greater levels of skill than ever before. At a time when the advent of the information super-highway is beginning to redefine the job functions of lower and mid-level workers in business and industry, it is vital that students complete a secondary school education in order to have better chances of finding employment. There is little reason for optimism in that respect as Mann (1986) reported that “A national estimate suggests that 25 percent of fifth graders will not make it through high school graduation” (Mann, p. 309).

There are many reasons for students dropping out of school, but Tanner (1990) suggests that “School based reasons are the most important self-reported explanation of dropping-out for all groups of adolescents” (p. 80). Chief among the school-based reasons is poor performance in school. When asked why they had dropped out of

---

6The data for this study were collected by Heather Gerrity.
school, more than one third of the students said “Because I had bad grades,” or “because I did not like school” (Mann, 1986, p. 309). These findings were substantially similar to those reported by Ekstrom, Goertz, Pollack, and Rock (1986). Therefore it seemed reasonable to assume that students who dropped out of high school would have lower metacognitive knowledge monitoring abilities than regular students. This expectation was examined in Study VI.

Participants and Procedures

The word list and vocabulary test employed in Studies II–V were administered, together with some test anxiety scales, described later in this chapter. The text passage was not used.

A total of 89 subjects participated. The dropout group consisted of 42 individuals (14 female) who had dropped out of high school and were attending a General Equivalency Diploma program. The continuing student group consisted of 47 students (16 female) who had a school GPA of at least B-. None of these students had given any indication that they were at risk of dropping out of school.

Results and Discussion

Two MANOVAs, identical to those of the preceding study, were computed to determine the significance of differences between the high school dropouts and continuing students. The first analysis found significant overall group differences (Transformed Wilks lambda $F(3,79) = 4.08, p > .01$, effect size = .134) in knowledge estimates (+ +, +-, and - - scores). Univariate analyses indicated that the dropout group (Mean = 12.7) differed from the continuing students (Mean = 16.2) on the + + scores ($F(1,81) = 8.83, p < .01$) and on the + - scores (dropout Mean = 10.6, continuing students 8.5; $F(1,81) = 6.11, p < .02$). A similar analysis of actual knowledge (+ + and + - scores) also indicated significant, though somewhat smaller group differences (Transformed Wilks lambda $F(2,80) = 4.61, p < .01$, effect size = .103). Univariate analyses indicated that only the + + score difference was significant.

The results indicate that, as expected, students who dropped out of high school had less effective knowledge monitoring abilities than did continuing students. Analysis of raw score differences yielded similar, though somewhat smaller effects. The results suggest that the poorer knowledge monitoring abilities of students who dropped out of school may have made school work more difficult for these students and contributed to poor performance, a picture that is consistent with the descriptions in the literature of school dropouts.
Summary: Knowledge Monitoring and School Learning

As expected, the four studies found significant relationships between metacognitive knowledge monitoring scores and school learning. The studies used different types of samples ranging from regular college students, those about to enter college and enrolled in a prefreshmen skills program, vocational and regular high school students, and those who dropped out of school. Because relationships with knowledge monitoring were in the expected direction for the different samples it may be inferred that the KMA has some generalizability across a variety of student groups. In most of the studies, the KMA scores accounted for more variance than raw vocabulary scores supporting the construct validity of the procedure.

PERFORMANCE EVALUATION AND METACOGNITIVE KNOWLEDGE MONITORING

The studies reported in the preceding section described the relationships between knowledge monitoring and school learning. The grades students receive are a function not only of their domain knowledge, but also of the standards and types of evaluations and grading procedures used by instructors. These factors potentially add error to the relationship between knowledge monitoring scores and GPA. In view of the fact that the KMA assesses students' abilities to estimate their knowledge, it was reasoned that students who were capable of accurately estimating their word knowledge on the KMA should also be more accurate in predicting their performance on examinations related to their present studies before they take them, and how well they performed on those examinations after they were completed. This section describes three studies examining these questions.

There has been some research on the prediction of performance in courses and on tests, though none of these related the predictions to metacognition or knowledge monitoring. Keefer (1971) found that college students who accurately estimated their performance achieved at a significantly higher level than less accurate estimators, and had a more positive self-concept than their low-estimating counterparts. Holen and Newhouse (1976) found that students' predictions of their grades on a course examination correlated as highly with actual performance as their GPA, and were significantly more accurate predictors than other variables, such as grades in prerequisite courses or GPA. Furthermore, students' performance predictions contributed significant
unique variance to predictions of actual final grade, above that contributed by high school and college GPA, or grades in prerequisite courses. Harris (1990) found that accurate estimators of test performance in psychology earned a significantly higher final average in introductory psychology than did low and less accurate estimators.

The research on performance estimation suggests that more capable students make more accurate predictions of their performance than their less able counterparts. Because the studies described in the preceding section found that accurate KMA scores were associated with higher GPA, the findings dealing with performance estimation support the rationale that students who make accurate metacognitive assessment of their knowledge should make more accurate predictions of test scores than would less accurate students.

Study VII. Performance Estimation and Predicting Standardized Test Scores

In some of the previous studies all students responded to the metacognitive procedure before and after reading the text passage. The results indicated that metacognitive estimates before students read the text passage were somewhat more highly related to their GPAs than those obtained after reading the passage, although the opposite findings emerged in studies relating knowledge monitoring to reading comprehension. A further purpose of this study was to vary the administration of the text passage in order to examine its contribution to students’ estimates of their test performance. Furthermore, it was considered useful to examine performance on a standardized test of known reliability to reduce possible error. Studies I and II used a standardized measure of reading comprehension as the criterion, and the results relating test performance to word KMA seemed more positive than the comparisons with the less reliable student grades. Therefore, the use of a test that had demonstrated reliability (.88) in a previous study seemed desirable.

It was expected that General Psychology students who could accurately monitor their knowledge would also be more accurate in predicting their actual and estimated scores on the Advanced Placement Test (AP) in Psychology (College Entrance Examination Board, 1988) before and after completing it, and that they would also earn higher scores on the test than their less accurate peers. Finally, as suggested by other studies of student’s estimation of their performance, it was predicted that they would expect to obtain higher grades in the course in which they were registered.
Participants and Procedures

The AP Examination in Psychology (College Entrance Examination Board, 1988) was administered to students enrolled in an introductory psychology class. Students received a description of the different areas covered by the AP test and were asked to predict how many of the 100 items they could answer correctly before the test was taken, and again after it was completed (postdiction). Half the sample \((N = 39)\) was randomly assigned to read the expository version of the text passage used in the two preceding studies before the word list, and the other half \((N = 38)\) received an unrelated task, the text selection titled "Teaching the Mentally Retarded" from the Sentence Verification Technique (Royer, Carlo, Dufresne, & Mestre, 1994), and then answered the questions on that passage. The same word list and vocabulary test used in Studies II–VI were then administered to all participants.

Students were also asked to predict their final grade in the Introductory Psychology class they were taking. On this campus the accuracy of their grade predictions could not be determined because regulations for the protection of students' privacy made it impossible to obtain information by which they could be identified.

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

Results and Discussion

More accurate metacognitive scores were expected for the group responding to the word list and vocabulary test after reading the text compared to the other group who received the SVT, which was irrelevant to the task. Surprisingly, MANOVA based on the total number of accurate estimates \([+ + \text{ and } - -]\) revealed no significant differences between the groups (see Figure 5). Examination of the basic eight scores \([+ +, + -, - +, - - \text{ for both explicitly and implicitly defined words}]\) indicated that there appeared to be some group differences (see Figure 6), but that these canceled each other out when the data were combined into total number of correct estimates.

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

Predictions of performance on the AP test were tested by splitting students at the median on total number of accurate metacognitive estimates [combining the ++ and --] and computing MANOVA to examine the significance of the differences on students’ pre- and postdictions of their AP scores, their actual AP score, and their expected final grade in the psychology class. There were no differences on the AP test data or on the expected final grades between groups who read either the text passage or the SVT \( F(4,69) p < 1 \). Differences between high and low metacognitive groups were significant \( F(4,69) = 2.83, p < .05; \) effect size = .141. Univariate tests indicated that the high knowledge monitoring group obtained higher AP scores (Mean = 43.6, \( F (1,72) = 7.81 p < .01 \)) than the low (Mean = 36.6), and that differences in expected final grade in the course just
failed of significance \( F(1,72) = 3.40, p < .10 \). There was no interaction between groups who read either the text or SVT and knowledge monitoring groups.

The data were also analyzed for the number correct on the vocabulary test by splitting the groups at the median on the number correct, and computing the significance of differences on the AP and final grade data. Those results were similar to the prior analysis using knowledge monitoring scores. That is, there were no differences between the groups who had read the text or received the SVT, and there was a significant difference between groups above and below the median on prescore \( F(4,69) = 6.47, p < .01; \text{ effect size } = .27 \). Univariate analysis again indicated only one significant difference on actual AP score between groups above (Mean = 45.4) and below (Mean = 34.2) the median on vocabulary score. Again there was no interaction among the variables. Unlike the prior studies, where differences in metacognitive estimates were usually greater than those on the vocabulary raw score, the effect size for these data using the vocabulary test results was larger than for the knowledge monitoring data (.27 compared to .14).

The results indicate that students high on vocabulary score and on the ability to monitor their word knowledge also obtained higher
scores on the AP exam and expected higher final grades in the course for which they were registered. The absence of group differences on predicted AP score before taking the test was not surprising because students were unfamiliar with the test, beyond being informed about the categories of knowledge covered. They had no information about the difficulty of the items, the types of preparation expected for the test, or specifically what they would be questioned on. The absence of differences on students' postdictions was a little more surprising because participants now had a much clearer idea about what the test covered. Perhaps this brief exposure to the test was inadequate to familiarize them with the domain covered by the AP test.

Study VIII: Knowledge Monitoring and Estimations of Academic Achievement

Ideally, of course, participants' performance estimates about both predicted and actual grades should have been studied in courses for which they were registered. Unlike the AP test, students should have enough information to make more reasonable predictions about their final grades in courses, based on their experience in the class, and with the subject matter, instructor, and procedures of the course. It was the purpose of this study to examine these expectations, in addition to attempting to replicate the findings for the AP data.

Participants and Procedures

The procedures were identical to those in the previous study with two exceptions. First, the predictions students made about their final grade were compared to the actual final grade obtained in the course. Second, students took 12 quizzes in this class (the instructor used the 10 highest quiz scores in the determination of the final grade) and the grades obtained on these quizzes were available as additional dependent variables.

A total of 75 students enrolled in the Introductory Psychology class participated in this study. The students received extra credit for taking part in the research.

Results and Discussion

The first set of analyses were computed to examine the consistency between the findings of this study and the preceding one. As in the prior study, a test for significance of differences between the group who read the text and the SVT on the + +, + -, and - - scores revealed no differences between the groups. When the component
scores, based on explicit and implicitly defined words, were examined, overall differences between the groups were significant \( F(6,68) = 2.57, p < .05 \). Univariate analysis indicated that the group reading the text had fewer \( - - \) scores for explicitly defined words \( (F(1,73) = 7.69, p < .01) \) and higher scores for the \( + + \) explicitly defined words \( (F(1,73) = 7.29, p < .01) \). These results are consistent with those of the preceding study and suggest that combining the data may obscure existing group differences. Both sets of results point to the importance of conducting a study specifically designed to determine which set of data are the best indicators of the latent knowledge monitoring variable.

The analysis of differences between high and low knowledge monitoring groups on predicted, postdicted, and actual AP scores, and final grades was also similar to that in the preceding study, with one addition—students’ actual final grades in the course were available as an additional dependent variable. Two groups were created by splitting students at the median on total number of accurate metacognitive estimates and computing a MANOVA to examine the significance of differences on the AP and grade data; nine students were eliminated due to missing information. No differences between groups who read either the text or the SVT were found \( (F(S,58) = 1.37) \). Unlike the prior study, the differences between metacognitive groups only approached significance \( (\text{Transformed Wilks lambda } F(S,58) = 2.21, p = .066; \text{ effect size } = .16) \). Univariate analysis indicated that the high metacognitive group had significantly higher AP scores \( (\text{Mean } = 45.2) \) than the low group \( (\text{Mean } = 36.7; F(1,62) = 10.02, p < .01) \); there were no differences on expected score either before or after the AP exam was taken, or on expected and actual final grades.

The findings that the high and low knowledge monitoring groups differed only on actual AP test performance, rather than on any of the estimates, also replicated those of the prior study. The failure to find differences on final grades may have been a function of the limited range of the grades; A–D grades (no F grades occurred in this sample) were converted to their numerical equivalents yielding only four scores. Furthermore, 76% of the grades were B or higher, further limiting their variability. The interaction between metacognitive group and those who read either the text or SVT was of borderline significance \( (F(5,58) = 2.18, p = .07) \), probably principally attributable to the fact that the low knowledge monitoring group’s estimates of their AP scores and their final course grades were actually higher than that of the high monitoring group, although their actual scores and grades were lower than those of the other students.
An identical MANOVA was computed with students split at the median on the vocabulary test as the independent variable. There were no differences between groups who read either the text or the SVT. There were significant differences between the groups (Transformed Wilks lambda $F(5,58) = 5.70, p < .001$; effect size = .33). Univariate analysis indicated that the high vocabulary group also had higher AP scores (Mean = 47.0, $F(1,62) = 22.89, p < .001$) than the low group (Mean = 35.1). Unlike the analysis based on metacognitive estimates, the high vocabulary group also received higher final grades (Mean = 90.4, $F(1,62) = 5.24, p < .05$) than the low group (Mean = 85). The interaction between groups who read either the text or SVT and vocabulary groups were not significant (Transformed Wilks lambda $F(5,58) = 2.12, p = .076$) even though the lower vocabulary group predicted and postdicted higher AP scores and final grades and actually obtained lower scores on all three.

The second set of analyses examined the relationship between the knowledge monitoring scores and in-class student performance indices, such as the quizzes administered to students and scores on the essay and multiple-choice parts of the final examination. Because the instructor informed students that only the 10 highest scores on the 12 quizzes would count for the final grade, many students missed some quizzes. Therefore, for students taking at least 10 of the quizzes, the mean score on all the quizzes taken was used as one of the dependent variables. Students were split at the median on the correct knowledge monitoring scores, and MANOVA was computed on the quiz and final examination data; missing data limited this analysis to 70 students. No significant differences on class performance indices were obtained between the groups taking the SVT or reading the text $F(4,63) = 1.04$). There was an overall significant difference between the high and low knowledge monitoring groups (Transformed Wilks lambda $F(3,64) = 4.36, p = <.01$, effect size = .17). Univariate analyses indicated that the high knowledge monitoring group had significantly higher scores on the multiple-choice part of the final examination (Mean = 25.1) compared to the low group (Mean = 21.2, $F(1,66) = 12.66, p < .01$). Differences between the groups on the mean quiz score were not significant ($F(1,66) = 3.02, p = .09$), although the high knowledge monitoring group received higher scores (Mean = 4.51—each quiz had a total of six raw score points) than the low group (Mean = 4.1). There was no interaction between knowledge monitoring and whether groups read the text or not.

The identical analysis was computed with students split at the median on the number of words correct on the vocabulary test as the
independent variable. The high and low vocabulary score groups had overall differences (Transformed Wilks lambda $F(3,64) = 6.44, p < .01$, effect size = .232). The high vocabulary group had significantly higher scores on both essay (Mean = 17.2, $F(1,66) = 7.44, p < .01$) and multiple-choice (Mean = 25.5, $F(1,66) = 18.72, p < .001$) parts of the final exam, and on the mean of the quizzes (Mean = 4.6, $F(1,66) = 7.13, p = .01$) than the lower scoring groups (Means = 14.5, 20.9, and 4.0 respectively). In this study as in the preceding one the differences between vocabulary score groups were greater than the metacognitive estimates for differentiating students on the AP and final grade data (.33 effect size vs. .16) and on the class tests (.23 compared to .17 effect size).

**Knowledge Monitoring and Performance Estimation Among Vocational High School Students**

In Study VI, examining relationship between knowledge monitoring and school learning among vocational high school students, the participants were also asked to predict and postdict their grades on a course final examination; the actual score on that test was available as a dependent measure. MANOVAs indicated that neither the metacognitive knowledge monitoring estimates, nor the raw scores were significantly related to either of the dependent variables. The failure to find any differences on actual score is at variance with the findings of the two preceding studies using college students.

There are a number of differences between the studies using vocational high school and college students, in addition to the population differences, that may account for the diverse findings. The vocational high school students were asked to predict performance on a final exam in the class they were taking, and presumably had a much better idea of the content of the exam and how to prepare for it than the college psychology students, who had very little basis for knowing what to expect on the AP test, and could not prepare for it at all. Furthermore, because the vocational students had been graded on other exams in that class, they—unlike the college students—knew what grade to expect from their prior history in that class. These prior experiences may have been more important in determining the vocational high school students' estimates than either their knowledge, or the metacognitive knowledge monitoring abilities.

**Summary: Performance Estimation and Knowledge Monitoring**

One striking finding of two of the studies using college students was that the strongest effects were found for students' actual perfor-
formance, either on tests or in class, rather than for their estimates. Students' estimated performance on the AP exam, or their predicted class achievement, were typically not significantly related to KMA scores. On the other hand, performance on the AP test, or in class final exams (at least the multiple-choice part of the exam in Study VIII) were significantly related to knowledge monitoring. These results may be partially attributable to unrealistic estimates of students in the lower knowledge monitoring groups that were often higher (though not significantly so) than those of students in the high knowledge monitoring group.

There is a large difference between the accuracy of vocational high school students' test performance estimates and that of college students. The correlations between predicted and actual scores for the vocational students were .71, and .75 \( (p < .001) \) for postdiction; comparable results for college students in Study VII were .13 and .16, both nonsignificant, and for Study VIII they were -.14 and -.12, also nonsignificant. The greater accuracy of the vocational students is probably attributable to their familiarity with the material they were tested on, compared to the novelty of the AP test for the two college samples. As expected, the relationships were higher, though not significantly so, for postdiction, when students knew what was covered on the test, than for predictions confirming findings by Pressley and Ghatala (1990) who reported that students were generally more accurate in predicting their recall of text after completing a test than before taking it.

In both of the studies using college samples, the analysis of school performance data based on actual knowledge (number correct on the vocabulary test) accounted for more variance than comparable analyses using knowledge monitoring scores. It seems possible that students' achievement in classes is best predicted by actual knowledge, rather than estimates of it. Furthermore, in view of the nonsignificant relationships for the vocational high school sample between either knowledge or metacognitive estimates and class final exams, it seems likely that domain specific knowledge may be most useful for such predictions.

An important question to investigate is whether knowledge estimates in the domain in which school instruction and evaluation are likely to occur account for more variance than the actual knowledge, or than either estimates or knowledge of fairly general materials such as those used in these studies. The prior research assumed that the word list, vocabulary, and text passage were similar to the kind of material students would learn in nontechnical areas of school instruction. The studies relating knowledge monitoring to school learning found KMA relationships with achievement in English and Hu-
manities courses, but not for Science and Social Science. These results suggest that general knowledge, or metacognitive estimates of that knowledge are less useful in more technical areas, which rely on a domain specific technical vocabulary, than they are in subjects that have a more widely shared knowledge base and vocabulary.

**METACOGNITION AND MATHEMATICS**

All of the studies described so far used the KMA in the domain of declarative word knowledge, and employed similar or identical versions of the materials. A question arises whether the procedure generalized to domains other than vocabulary, such as mathematics. Like vocabulary, mathematics is of special interest because it is also important in school learning. In addition, however, the computation and problem solving in mathematics involve more procedural knowledge than does vocabulary learning. Thus, one purpose of the two studies described below was to examine the applicability of the knowledge monitoring procedure to the domain of procedural knowledge in mathematics.

The research described above used relatively mature students, predominantly those attending college; only two investigations used high school students. A further question examined in the next two studies was whether the KMA was equally useful with younger, elementary school students.

**Study IX. Monitoring Mathematical Problem Solving Among Elementary School Students**

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

---

7The data for this study were collected by Dhalma Rosado. This investigation was presented as part of a paper at the annual convention of the American Educational Research Association, April 1995, San Francisco, CA.
Participants and Procedures

A list of 30 mathematical questions was constructed (20 computation, and 10 problem-solving items); the items were selected from the students’ fifth-grade mathematics curriculum. Students were first asked to take 6 minutes to determine if “you feel able to solve these problems. Do not solve them now,” giving them an average of 12 seconds per problem. During a later session, the same 30 questions were readministered, and students were given 40 minutes to actually solve the problems. A number of anxiety scales were also administered.

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

Results and Discussion

The scoring for the mathematics materials was similar to the vocabulary KMA. Four scores, like those used in the prior studies, were generated: Students felt that they could (a) solve a problem and did so (+ +); (b) not solve a problem and did not (- -); (c) solve a problem, but did not (+ -); and (d) not solve a problem, but did (- +). The results dealing with anxiety will be discussed later in the chapter.

There were no differences attributable to gender on students’ metacognitive estimates, so these data were pooled for further analysis. The knowledge monitoring scores were correlated with the total math score on the Metropolitan Achievement Test (1985) obtained from the students’ records. The correlations are displayed in Table 4. The last row in that table represents the number correct on the math test. The + + and - - scores were combined to indicate correct estimates of students’ ability to solve mathematical problems, and the - + and + - scores were added to form the incorrect estimates.

Table 4 indicates that three of the four estimates were significantly related to students’ mathematics achievement. The correlation between number correct on the math test and Metropolitan score was .52. When that relationship is compared to the correlation of .73 between Metropolitan score and + +, or the correlation of .76 between the Metropolitan score and total number of correct estimates, it is clear that metacognitive estimates of the ability to answer the questions are more substantially related to mathematical achievement than the number of problems solved correctly, irrespective of esti-
metacognitive knowledge monitoring. That finding was confirmed by regression analysis. When the number of correct estimates, incorrect estimates, and total number right were in the model, only the correct estimates contributed significantly to the prediction of Metropolitan score ($R^2$ Change = .08, $F (3,45) = 8.52, p < .01$). These results confirm the basic assumption of the knowledge monitoring procedure that students' metacognitive judgments contribute significant independent variance beyond those accounted for by number correct on a test.

The results support predictions regarding the relationships between the procedural KMA in mathematics and achievement in that domain. As expected, there were substantial correlations between students' estimates of their ability to solve mathematical problems and their achievement in mathematics. Also as expected, inaccurate assessments were negatively related to achievement. Although no causal inferences about mathematical achievement and knowledge monitoring can be made from these correlational data, the fact that the variables covary as expected supports the generalizability of the procedure and suggests that the technique is useful for further research in mathematics.

**Study X: Relationship of Procedural KMA in Mathematics with Age and Achievement**

The prior study provided encouraging evidence of the knowledge monitoring procedure's applicability to procedural knowledge in mathematics. Furthermore, the results of Study IX also indicated that

---

8The data for this study were collected by Audrey D'Agostino. The study was part of a paper presented at the annual convention of the American Educational Research Association, April 1994, in New Orleans, LA.
the KMA could be used with elementary school students. Because metacognition is often viewed as a developed ability and assumed to increase with age, one purpose of the next study was to investigate whether procedural knowledge monitoring in mathematics would also increase with age. The preceding study indicated a high relationship between KMA scores in mathematics and achievement test scores in that domain. Study X also examined whether knowledge monitoring scores were related to teachers' judgments of mathematical ability.

Participants and Procedures

Students were presented with 15 mathematical word problems involving addition and subtraction. The problems were set in the context of an ice cream store and students received a menu of prices for different products referred to by the word problems. The materials were prepared in two versions presumed to elicit varying interest levels among students. The results dealing with interest are discussed later in the chapter. The materials were administered on two days during regular class periods. On the first day, students examined the

![Graph showing Mathematical Knowledge Monitoring Scores by Grade Level.](image-url)
4. METACOGNITIVE KNOWLEDGE MONITORING

problems and estimated whether they could solve them or not; on the second day the students were asked to solve the problems.

Students (N = 164, 70 female) were selected from the fourth, fifth, and sixth grades of a school attended largely by children from minority groups. Mathematical ability was determined by teachers' judgment; 59 students were classified as being in the low, 67 in the medium, and 81 in the high ability groups.

Results and Discussion

Students' responses were assigned a score of 1 for correct estimates (combining the + + and - - scores), and 0 for incorrect estimates (combining the + - and - + scores). Due to a computer malfunction, the raw data were not available for rescoring to form the same scores used in the other studies. The data were then submitted to a 3 (grades) x 2 (sex) x 2 (group) x 3 (math ability) analysis of variance. As indicated above, the results dealing with interest are discussed later.

As expected, a significant increase in knowledge monitoring scores from grades 4 to 6 was found ($F = 34.66$, df = 2, 144, $p < .001$, $\eta^2 = .26$; see Figure 7 for plot of the data). Also as expected, the results indicated that knowledge monitoring scores increased with math-

![Figure 8. Mathematical Knowledge Monitoring Scores by Math Achievement Group.](image-url)
ematical ability \((F = 15.25, df = 2, 144, p < .001, \eta^2 = .18; \) see Figure 8 for plot of the data). These results offer further support for the construct validity of the knowledge monitoring procedure because older or more capable students were expected to have better metacognition than their younger, less capable counterparts. There were no significant differences attributable to sex.

**Summary: Knowledge Monitoring and Mathematics**

The results of the two studies using the procedural KMA in mathematics were quite positive regarding its applicability to that domain. The relationship of monitoring scores to achievement in mathematics in Study X are similar to the correlations with math achievement test scores reported in Study IX, and both indicate strong relationships between metacognitive knowledge monitoring and achievement in mathematics. The increases in metacognitive ability associated with age reported in Study X also support that relationship. Furthermore, 10 of the studies reviewed in this chapter examined students’ estimates of their declarative word knowledge. Because most of the items in both mathematical studies were composed of procedural knowledge needed to solve word problems, the results suggest that the KMA may be applicable to that type of knowledge as well.

**METACOGNITION AND AFFECT**

The paradigm shift to a cognitive orientation in psychology generated a great deal of research to clarify the cognitive processes controlling learning from instruction. Although that work has identified many cognitive processes that are important in human learning, the impact of affective processes on such learning has received considerably less attention (Tobias, 1992, 1994a, b). The aim of the research discussed in this section is to forge a link between affect and cognition by examining the influence of affective variables such as anxiety and interest on metacognitive knowledge monitoring.

**The Impact of Anxiety on Knowledge Monitoring**

One of the affective variables that has been the subject of a great deal of research, both in educationally relevant situations and in others, has been anxiety and its impact on learning. In general, that research has suggested a negative relationship between different forms of anxiety and learning from instruction (Tobias, 1992; Hembree, 1988). It has been suggested (Tobias, 1985, 1992) that interference in
students’ performance as a result of anxiety was attributable to reduced cognitive capacity available for task solution. It was reasoned that the central representation of anxiety absorbs some proportion of cognitive capacity, leaving a reduced amount available for allocation to work on tasks. The further absorption of capacity required by an executive process such as metacognitive knowledge monitoring was expected to be especially debilitating for highly anxious students whose cognitive capacity is expected to be reduced by students’ concerns about their test anxiety. Therefore, a negative relationship between anxiety and knowledge monitoring was anticipated because “highly test anxious students can be expected to have less adequate metacognitive abilities than those with lower anxiety” (Tobias, 1992, p. 28).

Knowledge Monitoring, Reading Comprehension, and Test Anxiety

It will be recalled that Study II examined the relationship of the knowledge monitoring procedure to anxiety, in addition to reading comprehension. The worry subscale of the Test Anxiety Inventory (Spielberger et al., 1980) was administered to 117 undergraduates from a large urban university; 65% were women.

As expected, the more highly anxious participants also performed less well on the KMA. Those lower in anxiety achieved a significantly higher number of “hits” than those prone to higher levels of anxious worry ($t(115) = 4.92, p < .001$), and in general the less anxious had higher levels of metacognitive word knowledge as measured by $d'$ (multiple $r$ squared ($t(115) = 4.07, p < .001$), confirming the expected negative relationships between knowledge monitoring and test anxiety.

Knowledge Monitoring in Mathematics and Anxiety

Study II found the expected negative relationship between knowledge monitoring procedure scores and anxiety in the vocabulary domain. Study IX, in addition to investigating the extension of the knowledge monitoring procedure to mathematics, also studied its relationship with both test and mathematics anxiety.

As part of Study IX the Fenema-Sherman (1976) scales assessing math anxiety and attitudes towards mathematics were administered to the 51 participants (see the earlier description of Study IX) in a first session. In order to assure that the participating elementary school students could understand the questions, each item was read aloud as students read the materials. The Worry-Emotionality scale (Morris,
Davis, & Hutchings, 1981), a 10-item Likert-type measure of these components of state test anxiety, was also administered. Students’ mathematical achievement was determined from their scores on the Metropolitan Achievement Test (1985) obtained from school files.

In Study IX no sex differences in the effects of anxiety were found, so the data for all students were pooled. The relationships between knowledge monitoring and mathematics anxiety (scored in the direction of higher anxiety yielding higher scores) and attitudes towards mathematics, as well as with worry and emotionality are shown in Table 5.

Table 5. Correlations Between Knowledge Monitoring Scores and Anxiety in Mathematics.

<table>
<thead>
<tr>
<th>Score</th>
<th>Math Anxiety</th>
<th>Worry and Emotionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>++</td>
<td>-.42**</td>
<td>-.22</td>
</tr>
<tr>
<td>+</td>
<td>.32*</td>
<td>.25</td>
</tr>
<tr>
<td>- +</td>
<td>.38**</td>
<td>.23</td>
</tr>
<tr>
<td>- -</td>
<td>.00</td>
<td>.20</td>
</tr>
<tr>
<td>++ and - -</td>
<td>-.46**</td>
<td>-.15</td>
</tr>
<tr>
<td>- + and + -</td>
<td>.46**</td>
<td>.33*</td>
</tr>
</tbody>
</table>

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

Table 5 indicates that, as expected, mathematics anxiety was negatively related to incorrect metacognitive estimates and positively related to correct ones. The correlation between number right and math anxiety was -.25 and not significant, though the relationships with metacognitive estimates were negatively and significantly related to test anxiety. The negative relationships between metacognition and anxiety are generally similar to those found in Study II, confirming expectations that anxious students have lower metacognition than their less anxious colleagues.

Anxiety and Knowledge Monitoring Among High School Dropouts and Continuing Students

Study VI investigated whether knowledge monitoring differed between continuing students and high school dropouts. An addi-
tional purpose of that study was to examine the differences in anxiety between high school dropouts and continuing students, as well as the relationship between anxiety and metacognitive knowledge monitoring. In this study the Test Anxiety Inventory (Spielberger et al., 1980) was given to all participants, followed by two administrations of the Worry-Emotionality Scale (Morris et al., 1981). Initially, participants completed the Worry-Emotionality scale in terms of the way they felt while being tested in general; when the scale was re-administered after the vocabulary test, students were asked to respond the way they felt while completing the vocabulary test. It will be recalled that the dropout group consisted of 42 individuals (14 female) who had dropped out of high school and the 47 (16 female) continuing students who had a school GPA of at least B-.

Surprisingly, the results of a MANOVA indicated that there were no anxiety differences between high school dropouts and continuing students on any of the seven anxiety scores (the three Test Anxiety Inventory scores: Worry, Emotionality, and Total, in addition to four Worry and Emotionality scores from each administration of those scales). That finding is puzzling in view of the reports in the literature that poor performance in school, and presumably on tests, was a major reason for students dropping out of high school. One reason for these findings may rest with the problems to which self-report measures in general, and self-reports of test anxiety in particular, are subject. It is easily possible for students to minimize or deny responses indicative of test anxiety on these measures, and to present themselves as not caring about how well they might function on tests. The knowledge monitoring procedure, however, made it difficult for students to present themselves in a more favorable light, and that may account for the findings of group differences in metacognitive knowledge monitoring and their absence in test anxiety.

Most of the zero order correlations between KMA scores and anxiety indices were negative, and a fair number were significant. Multiple linear regression analyses were computed with the KMA scores as the dependent variable, and the anxiety scores as the independent variable. Results indicated that the anxiety scales had a significant impact only on the + + scores, $R^2 = .25$, $(F(7,72) = 3.43, p<.01)$; significant beta weights were found for Emotionality, on the Worry-Emotionality Scale taken after students completed the vocabulary test ($t = 2.74$). The regression analysis also indicated that none of the other KMA scores were significantly related to the anxiety scales. In view of the number of anxiety and knowledge monitoring scores, the findings of significant relationships for some of them is not
surprising. In general, however, the results of this study suggested that there was little association between metacognitive knowledge monitoring and anxiety.

Anxiety and Knowledge Monitoring Among Vocational High School Students

A further purpose of Study V, examining metacognition among Vocational High School students, was to study the relationship between anxiety and metacognitive knowledge monitoring, as well as between anxiety and achievement. In addition to the rationale relating metacognition to anxiety, it was also expected that students with lower GPAs should have higher anxiety than those who learned more effectively. These two questions were examined in this study.

Recall that the anxiety scales, and the order in which they were administered in Study V, were identical to those employed in the study of high school dropouts (Study VI). The Worry-Emotionality scale (Morris et al., 1981) was administered first and students were asked to respond in terms of the way they felt while taking tests in general. The Test Anxiety Inventory (Spielberger et al., 1980) was then given, followed by a second administration of the Worry-Emotionality scale with instructions for students to respond the way they felt while taking the vocabulary test. A total of 61 students (59 male) participated in this study.

The significance of differences in anxiety scores between the participants in Study V above and below the median GPA in vocational high school was examined by computing a MANOVA. Surprisingly, there were no differences between the GPA groups on any of the seven anxiety scores. Also, much as in Study VI, most of the zero order correlations between knowledge monitoring and anxiety were negative. Multiple linear regression analysis was then computed with the metacognitive knowledge monitoring scores as the dependent variable, and the anxiety scores as the dependent variable. None of the regression equations were significant for this sample.

Summary: Anxiety and Metacognitive Knowledge Monitoring

The evidence regarding the relationship between anxiety and metacognitive knowledge monitoring is mixed. Significant negative relationships were expected and found in two of the studies, one in mathematics and the other using vocabulary. On the other hand, two other studies failed to find any evidence of differences. Study II, in which significant negative relationships with anxiety using the vo-
cableary materials were found, had a much larger sample than the studies using vocational high school students (see Study V) or high school dropouts (see Study VI). Because many of the test anxiety-metacognitive knowledge monitoring relationships in the two latter studies were, as expected, in the negative direction, and because some of the regression analyses between these variables approached significance, further research with larger samples is clearly needed to clarify the relationship between anxiety and knowledge monitoring. The results of Study II suggest that metacognitive word knowledge and test anxiety each contributed to performance on less challenging reading. On demanding material, however, test anxiety and metacognitive knowledge monitoring ability appear to interact to affect performance. The highly anxious examinee, regardless of metacognitive ability, performed less well on the more demanding reading tasks, suggesting that anxious worrying can interfere with strategic use of metacognitive skill when the performance tasks are cognitively demanding. That finding is in accord with the anxiety-cognitive capacity model (Tobias, 1992), because more demanding tasks require greater proportions of cognitive capacity that may not be available as a result of the resources absorbed by anxiety. Further research is required to pursue that intriguing finding.

In Studies V and VI, the failure of a number of anxiety indices to differentiate between either high school dropouts and continuing students, or between students above and below the median in GPA was surprising. A meta-analysis of 562 studies dealing with test anxiety (Hembree, 1988) had indicated that lower achieving students had higher test anxiety than their more capable counterparts. Although there had been no prior research specifically relating test anxiety to dropping out of high school, the bulk of this literature has indicated that the concern of students about their academic achievement was a major factor in dropping out of school, clearly suggesting that differences in test anxiety could be expected. As mentioned above, the fact that both the studies dealing with dropouts and vocational high school students found significant knowledge monitoring differences in the expected direction, and neither found differences on a group of seven test anxiety scales re-emphasizes some of the problems with self-report measures described at the beginning of this chapter.

Although the nonsignificant results for anxiety in Studies V and VI may be attributable to small samples, or to other unknown factors, it should also be noted that the tendency of participants to present
themselves in a more positive light may well have contributed to the nonsignificant findings. One advantage of the KMA is that, because students do not report on either their feelings or their cognitive processes, it is difficult for them to present themselves more favorably. Of course, students could easily claim to know more words than they actually do. However, that claim is immediately checked by the administration of the vocabulary test making it harder for students to appear in a more positive light.

Knowledge Monitoring and Interest

There has been a good deal of recent research on the effects of interest on learning (Renninger, Hidi, & Krapp, 1992). One reason for that increase is suggested in one definition: "intrinsically motivated behaviors are those the person undertakes out of interest" (Deci & Ryan, 1991, p. 241). From that perspective, clarification of the effects of interest also adds to an understanding of the impact of intrinsic motivation on learning. Second, interests appear stable and long lasting among adults (Hidi, 1990; Schiefele, 1991), suggesting that instruction adapted to students' interests may have positive motivational characteristics for long periods of time. Third, interests are ubiquitous—everyone is interested in something. Fourth, findings of surprisingly variable and ineffective cognitive processing of instruction (Paris, 1988; Tobias, 1989) suggested that these results may be attributable to the possibility that students' interests or motivation were not engaged by the materials used in such studies. Finally, research on interest provides a useful and educationally relevant avenue for study of the relationship between affect and cognition (Tobias, 1989, 1994a, b)—a much needed clarification in order to obtain a more complete picture of people's day to day functioning.

Research has indicated that reading comprehension and recall are facilitated when students work on material related to their interests (Renninger et al., 1992). Furthermore, Schiefele (1990, 1991, 1992a, b) found that comprehension of interesting text was "deeper" (i.e., more likely to be propositional than verbatim). Little is known, however, about the cognitive processes that mediate the effect of interest on comprehension and recall of reading. Therefore, it was recommended (Renninger et al., 1992; Tobias, 1994a) that research concentrate on the identification of the processes invoked by interest to facilitate learning. The purposes of the studies reported in this section of the chapter are to examine whether interest improved students' metacognitive knowledge monitoring.
Situational Interest, Topic Interest, and Knowledge Monitoring

Two types of interest, situational and topic, have been distinguished (Renninger, Hidi, & Krapp, 1992). Situational interest is elicited by aspects of a situation, such as its novelty or intensity, and by the presence of human interest factors contributing to the attractiveness of different types of content. Topic interest refers to peoples’ relatively enduring preferences for different topics, tasks, or contexts and how they influence learning. The effects of both types of interest on metacognitive knowledge monitoring were investigated in this study. It was expected that students with greater topic interest, and text that elicited situational interest, would lead to more accurate knowledge monitoring. Furthermore, because interest has been found to lead to deeper types of text processing (Schiefele, 1990, 1991, 1992a, b), it was expected that students would make more accurate knowledge monitoring judgments on words requiring intense processing if the material were interesting rather than neutral. The meanings of implicitly defined words must be inferred, whereas those defined explicitly merely require recall of the definitions. Therefore, it was reasoned that the meanings of implicitly defined words should be estimated more accurately on interesting content compared to more neutral content.

Study III Revisited

Recall that two groups of students were used in Study III, nursing students and college freshmen. Because the passage dealt with heart disease, it was expected that nursing students would have greater topic interest in that material than would the freshmen. Situational interest was varied by converting the expository passage to a narrative format. The narrative passage contained story attributes, such as character identification and life themes, which according to Hidi and Anderson (1992), increased the situational interest of passages. A principal character was introduced in the narrative version, which then described his efforts to learn more about coronary disease because his father had developed a mild form of that illness. The passage indicated that he was trying to help his father prevent the development of more serious coronary problems. This structure made it possible to include in the narrative version all the factual information present in the expository version of the passage. Only 84

---

9This study conducted by Sigmund Tobias was published in the Journal of Educational Psychology, 1995. See References.
of the 139 students completed all the materials during two sessions. Complete data were available for 33 nursing students and 51 freshmen.

In Study III the correct metacognitive judgments (combining + + and -- scores) were submitted to analysis of variance, with the correct estimates on explicitly and implicitly defined words—the dependent variables—treated as a repeated measure. In view of the importance of controlling for prior knowledge differences in interest research (Tobias, 1994a), students’ scores on the first administration of the vocabulary test were used as a covariate because the nursing students were more familiar with the heart disease material (Prescore Mean = 27.4, SD = 4.0) than were the freshmen (Prescore Mean = 20.1, SD = 5.3). Because there were an unequal number of females in the groups (24 of 51 freshmen and 28 of 33 nursing students), gender was added as a factor in the design. Thus, the ANOVA consisted of a full 2 (freshmen vs. nursing students) x 2 (expository vs. narrative passages) x 2 (gender) factorial design, with prescore as a covariate. Again, the two-level repeated measure consisted of the number of correct judgments on explicitly and implicitly defined words after reading the text. The main effect of the repeated measure was assessed in the “deviation” manner described by Delaney and Maxwell (1981).

The ANOVA results indicated that there was a significant overall difference between the freshmen and nurses \((F(1,75) = 4.99, p < .05)\), favoring nursing students. In addition, the mean number of correct estimates was higher for explicitly than for implicitly defined words \((F(1,75) = 8.27, p < .01)\). None of the other main effects or interactions was significant. The covariate, number correct on the first administration of the vocabulary test, exerted a significant effect on the dependent measures \((F(1,75) = 17.01, p < .001)\). The adjusted means for freshmen on correct estimates for explicitly and implicitly defined words were 13.7 and 12.5, respectively, and for nursing students the corresponding means were 15.0 and 14.1. Ideally, future research should use participants with similar prescores who differ in their interest in such medically relevant materials.

These results support the general hypothesis of enhanced metacognition for topic interest. As anticipated, nursing students, for whom the heart disease passage was more interesting than for freshmen, made more accurate metacognitive assessments of their vocabulary knowledge than the freshmen, even when differences in their prior knowledge of the vocabulary was controlled statistically. The expected differences attributable to situational interest were not found
because the narrative and expository passages resulted in similar KMA scores. Finally, contrary to expectations, explicitly defined words were judged more accurately than those that were implicitly defined for both nursing students and freshmen.

The absence of knowledge monitoring differences due to situational interest may be a function of the similarities between the expository and narrative texts. Even though the passage was altered to elicit differences in situational interest, ratings of interest on a Likert-type scale, in the original study and on a follow-up, failed to find any differences between the passages. Perhaps, greater differences in content are needed to result in situational interest differences than occurred in Study III.

**Interest and Knowledge Monitoring in Mathematics Among Elementary School Students**

Study X found that metacognitive knowledge monitoring ability in mathematics increased with grade and mathematical ability. A further purpose of that study was to examine the impact of personalizing instruction on metacognition. Research (Anand & Ross, 1987; Bracken, 1982; Herndon, 1987; Lopez, 1999, 1990; Ross & Anand, 1987; Wright & Wright, 1986) has shown that personalizing mathematical word problems by including materials such as the names of students, their friends’ and teachers’ names, or including materials related to their interests improved performance and attitudes to the materials. These, and similar, studies suggested that heightened interest was aroused by personalizing word problems. It was, therefore, hypothesized that the elevated interest should improve students’ metacognitive knowledge monitoring.

Participants in Study X (N = 164, 70 females, and all of whom were selected from fourth, fifth, and sixth grades of a school attended largely by children from minority groups) were randomly assigned to either interesting or control materials. In the interesting condition the names of classmates and teachers were included in word problems, whereas the materials used for the control group used standard rather than familiar names. In each condition, 15 mathematical word problems set in the context of an ice cream store were presented and students received a menu of prices for different products and were required to add and subtract menu items. A 12-item Likert-type scale designed to assess interest in the materials was also administered.

In this study students’ responses were assigned a score of 1 for correct estimates and 0 for incorrect estimates. The data were then submitted to a 3 (grades) x 2 (sex) x 2 (interest group) x 3 (math ability)
analysis of variance. The findings dealing with knowledge monitoring and mathematical ability and grade level were reported previously. There were no significant differences attributable to sex or to interest. However, there was an interaction between math achievement level, as determined by teacher judgment, and interest \((F = 6.02, \text{df} = 2, 144, p < .01, \eta^2 = .05; \text{see Figure 9 for a plot of the data})\).

The interaction, unlike the main effect found in the prior interest study, suggests that the personalization improved the performance of low ability math students, but had little effect on the other groups. It seems possible that setting the word problems in the context of an ice cream store may have raised the interest level of the materials for both groups, thus leading to the insignificant main effect for interest. In view of the known difficulties students have with math word problems (NAEP, 1979), it was thought to be important to make the materials interesting for both groups by situating them in an ice cream parlor. There is evidence that this setting did arouse the interests of all students. There were no differences \((F < 1.0)\) between the high and low interest groups on the 12-item Likert-type scale administered after

![Figure 9. Mathematics Knowledge Monitoring Scores by Interest Group.](image-url)
students completed the problem solving. Furthermore, there were also no differences between the high and low interest groups in the number of problems solved correctly. These findings indicate that even the low interest group may have found the materials more attractive than the math word problems usually received in school, and suggests that an overall facilitative effect for interest may be found when the materials elicit greater differences in interest between the groups.

Summary: Affect and Knowledge Monitoring

The findings of the anxiety and interest studies indicate that anxiety generally seems to have a negative effect on metacognitive knowledge monitoring, and that working on interesting materials seems to facilitate it. Further research is needed to answer many questions before these tentative conclusions can be stated with greater confidence. It seems, however, that the knowledge monitoring procedure is a useful way for studying the effects of affect on metacognition, and especially of investigating the effects of interest. There are a number of persuasive models specifying the cognitive processes mediating the impact of anxiety on learning (Sarason, 1987; Eysenck, 1988; Tobias, 1992). However, little is known about the cognitive processes by which such "positive" affective variables as interest and motivation facilitate learning. The knowledge monitoring procedure seems useful for further research relating metacognition to such positive variables as interest or intrinsic motivation.

METACOGNITIVE KNOWLEDGE MONITORING AND OTHER VARIABLES

Most of the studies reported above related the knowledge monitoring procedure to relatively traditional variables such as achievement in school, estimates of achievement, anxiety, and interest. Two studies examined the relationship of the knowledge monitoring procedure to other variables such as need for feedback and the procedure's ability to differentiate between different types of students. These studies are summarized below.

Study XI. Knowledge Monitoring Procedure and Need for Feedback\(^\text{10}\)

Feedback or reinforcement is one of the best known variables in learning research. Numerous studies have demonstrated that feedback facilitates learning. McKeachie (1974) suggested that the effects

\(^{10}\)The data for this study were collected by Nadia Seignon.
of feedback or reinforcement on human learning are not uniform, but may vary with individuals and situations. Ashford and Cummings (1983) found that the importance of feedback varied with individuals’ uncertainty and Tuckman and Sexton (1992) found that students in a no-feedback condition who held high beliefs in their own performance capability outperformed those receiving feedback, whereas the reverse was true for students of middle and low self-perceived capability. These results clearly supported the idea that there were individual differences in the need for feedback.

It was expected that the need for feedback ought to depend on students’ metacognitive capability to monitor their knowledge gathering activities. In an analysis similar to that proposed by Butler and Winne (1995), it was reasoned that students with accurate knowledge monitoring abilities probably rely more frequently on their own, or internal, feedback regarding the accuracy of their responses than on their less accurate peers. Such students are likely to have learned from experience that external feedback often duplicates the information supplied internally, and should require less externally supplied feedback than colleagues with less accurate knowledge monitoring abilities. Therefore, when students have a choice of whether they choose to obtain feedback or not, a negative relationship between KMA accuracy scores and amount of feedback was expected.

Participants and Procedures

A list of 25 words, appropriate for fifth grade students, and a vocabulary test based on the same words were developed. Participants were also given a reading test consisting of 11 narrative stories with an average length of 140 words, or 15 sentences. Each story had a blank to be filled in, and students were instructed to select a word from four choices appearing in the right margin for each blank; the words on the reading test and word list were different. Participants were told that the correct answer to each question was printed in the left margin of each page, covered by a tab, and they could look at the answers whenever they wished simply by lifting the tab. Participants were tested individually, and the number of times the tabs were lifted to inspect the correct answer were recorded.

A sample of 59 fifth grade students (35 females) participated in this study. The school was attended primarily by minority students.

Results and Discussion

Students’ need for feedback was operationally defined as the number of times they lifted the tabs covering the correct answers to
questions in order to check on the correctness of their response. The knowledge monitoring procedure was scored to determine students' accuracy in estimating their word knowledge and the results were then correlated with amount of feedback sought on the reading test. The results of that analysis are shown in Table 6.

As expected, the results indicated that amount of feedback was substantially related to the ability to monitor knowledge accurately.

Table 6. Correlations with Amount of Feedback.

<table>
<thead>
<tr>
<th>Score</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>++</td>
<td>-.50**</td>
</tr>
<tr>
<td>+ -</td>
<td>.38*</td>
</tr>
<tr>
<td>- +</td>
<td>.56**</td>
</tr>
<tr>
<td>- -</td>
<td>-.13</td>
</tr>
<tr>
<td>++ &amp; - -</td>
<td>-.79</td>
</tr>
<tr>
<td>- + &amp; + -</td>
<td>.76</td>
</tr>
<tr>
<td>Score (++ &amp; - +)</td>
<td>-.19</td>
</tr>
<tr>
<td>R</td>
<td>.84</td>
</tr>
<tr>
<td>R²</td>
<td>.71</td>
</tr>
</tbody>
</table>

The accuracy of knowledge monitoring was substantially and negatively related to amount of feedback ($r = -.79, p < .001$), as were the number of inaccurate estimates ($r = .76, p < .001$). Equally interesting was the finding that vocabulary knowledge, determined by the number correct on the vocabulary test, was not related to amount of feedback ($r = -.19$). The findings suggest that, as expected, need for feedback was heavily related to the ability to monitor one's knowledge accurately. Furthermore, estimates of students' knowledge were clearly the major contributor to that relationship, because actual knowledge was unrelated to amount of feedback.

An equally important aspect of this study and its results was the fact that a new word list and vocabulary test was developed, different from the materials used in any of the other studies described in this chapter. Therefore, the findings also indicated that the knowledge monitoring procedure has some generality across different types of vocabulary materials. Furthermore, this was the first study using a declarative vocabulary KMA with elementary school students, and the results suggest that the procedure was as applicable to younger students as were the mathematical materials used in Studies IX and X.
Study XII. Knowledge Monitoring Differences Among Learning Disabled and Hyperactive Students

It has been shown (Brown & Campione, 1986; Swanson & Trahan, 1992) that students diagnosed as learning disabled (LD) have lower metacognition than those who are diagnosed as not being LD. Similarly, students with Attention Deficit Hyperactivity Disorders (ADHD) have been succinctly described by Douglas, Barr, O’Neil, and Britton (1986) as having a cognitive deficit featured by an inability to stop, look, listen, and think. A review of research dealing with ADHD (Westby & Cutler, 1994) indicates that such students tend to have less effective complex problem-solving strategies and organizational skills, that they use less efficient strategies on memory tasks, that they “demonstrated deficits on all measures of study behavior. They studied for less time, expended less effort, and used poorer strategies....students with ADHD have significant deficits in executive processes” (Westby & Cutler, 1994, pp. 63-64). These deficits clearly suggest that ADHD students have less effective metacognition. Therefore, students diagnosed as LD, or ADHD, should have less accurate knowledge monitoring capabilities than students not affected by these conditions. It was the purpose of this study to test that hypothesis.

Participants and Procedures

A list of 35 words, and a vocabulary test based on the same words, were developed from the high school curriculum. Participants ($N = 90$) were selected from the ninth ($N = 29$) and tenth ($N = 61$) grades of a public high school in an urban area; there were 28 females and 62 males. LD and ADHD groups ($N = 30$ each) were formed by selecting students diagnosed by a school-based support team consisting of an educational evaluator, a school psychologist, and a social worker; scores on the Degrees of Reading Power (Touchstone, 1991) test placed these groups in the 15th percentile of the population. A contrasting student group ($N = 30$) was selected on the basis of having average reading ability on the DRP and no histories of special educational needs.

Results and Discussion

Three of the KMA scores (+ +, + - , and - -) were analyzed by MANOVA (the fourth score [- +] could not be entered due to linear

---

11The data for this study were collected by Julie Wilson.
dependencies), with sex and group as the independent variables. A significant overall difference among the groups was found (Transformed Wilks lambda $F(6,164) = 5.95$, $p < .001$, effect size = .179). Univariate analyses indicated significant differences among the groups on $++$ scores ($F(2,84) = 16.02$, $p < .001$; Control Group Mean = 28.4; LD Mean = 22.2; and ADHD Mean = 23.0). Univariate analyses also indicated another difference on the $- -$ score ($F(2,84) = 5.32$, $p < .01$; Control Group Mean = 1.5; LD Mean = 3.6; and ADHD Mean = 4.3); students in the control group had lower scores because they had fewer incorrect answers. There were no differences attributable to sex, and no interaction between sex and group was found.

A similar analysis of the number correct on the vocabulary test ($++$ and $- -$) also indicated significant group differences ($F(4,166) = 7.55$, $p < .001$, effect size = .154). Univariate analysis indicated that only the differences on the $++$ scores were significant; the group means are the same as for the preceding analysis. The results confirm expectations regarding differences between regular, LD, and ADHD students with respect to their ability to monitor their knowledge and differentiate between what they know and do not know in this domain. Although the results were similar when the dependent variable consisted only of the number correct on the vocabulary, the effect size on the latter analysis was smaller (.154 compared to .179). As expected, the control group of students without special needs were more able to differentiate between the $++$ and $- -$ words than students in the other two groups.

There were large differences in reading ability between the groups which may also have accounted for the group differences, irrespective of diagnostic category. It is often difficult to separate the effects of reading ability from research comparing LD, ADHD, and more traditional students because the presence of reading problems is one of the defining characteristics of the two former groups. Further research with more similar groups may clarify this problem. In any event, these results provide additional support for the construct validity of the metacognitive knowledge monitoring procedure. In view of the fact that this study, like the prior one, also developed a new list of words and vocabulary test, the results also support the generality of the procedure across different types of content.

GENERAL DISCUSSION

The findings of the 12 studies summarized above support the construct validity of the KMA. Comparable results were found for samples from student populations such as students in elementary
school, students attending regular and vocational high schools (including students diagnosed as LD and ADHD), individuals who dropped out of high school, and students in pre-college-admission status, and students in their second or higher year in college. Furthermore, substantially similar results were obtained for procedural knowledge in mathematics, in addition to declarative vocabulary knowledge based on three different vocabulary sets developed to be appropriate for elementary school through college levels.

The results suggest that the procedure has some generality across different populations, declarative and procedural knowledge, as well as different types of vocabulary. In view of the fact that the KMA may be group administered and/or given by computer, and is objectively scored, it seems to be a promising approach for the assessment of the knowledge monitoring component of metacognition. Furthermore, Studies V and VI indicated that the KMA made it less likely that students presented themselves in a more favorable light than self-report scales of anxiety, one of the problems inherent in self-report instruments. Although no data comparing the KMA to other metacognitive scales have so far been collected, we expect that this measure of knowledge monitoring is likely to be more accurate than self-report scales because students are less able to present themselves in socially desirable ways. It remains for further research to investigate this possibility.

The KMA's relationships with external criteria were somewhat variable. Relationships with standardized achievement tests were substantial and significant. For example, in Study I correlations with a reading comprehension test were .67. Similarly, relationships with achievement in mathematics were also substantial in Study IX ($r = .76$) and in Study X, significant effects were found for KMA differences in students' math achievement ($\eta^2 = .26$) and for higher levels of mathematical performance across three elementary school grades ($\eta^2 = .18$). Pintrich et al. (this volume) cite some of these findings as being among the most positive relationships between any metacognitive measure and external criteria. Relationships with need for feedback in Study IX were also found to be substantial ($r = .62$). Significant, though somewhat more moderate, relationships were found in studies in which the KMA differentiated between known groups such as regular students and dropouts (Study VI), or among LD, ADHD, and students without special needs (Study XII). Generally the lowest, though frequently significant, relationships were found between KMA scores and college grades. Presumably, as indicated previously, the low reliability of such grades accounts for the modest associations with
grades. It should also be noted that differences between the effects of knowledge estimates and actual knowledge, discussed below, should be considered in examining the effects.

A number of issues raised by the results require further research. These include such questions as: Do multiple administrations of the knowledge monitoring procedure increase its relationship with other variables? Which of the different scores are optimal indicators of knowledge monitoring abilities? Do estimates of knowledge account for more variance than the actual knowledge? These questions are addressed below.

KMA and Dynamic Assessment

Some of the studies described above administered the text passage to only a part of the sample, others did not use the text passage at all, and still others gave a word list and vocabulary test before and after students read a text passage from which the word meanings could be inferred. A question arises about the value of interspersing the text passage between administrations of the word list and vocabulary. Giving students a chance to update their knowledge has some similarities to dynamic assessment approaches (see Carlson & Wiedl, 1992; Guthke, 1992; Lidz, 1992) in which students are given the opportunity for new learning before being tested. Dynamic assessment procedures usually also include some intervention in students' attempts to learn, observations of their reaction to the intervention, and an evaluation of students' responses to the assistance as part of the assessment. Reviews have suggested (Carlson & Wiedl, 1992) that students' attempts to verbalize learning difficulties, and receiving elaborated feedback about their efforts, contribute heavily to the value of dynamic assessment. The KMA differs from dynamic assessment procedures because it does not include any of these additional attempts to facilitate learning; students are merely given another opportunity to learn the words from a text passage without any other assistance.

The results of the present research indicate that the opportunity to learn the meanings of some words from the text was most important only in the first study relating the knowledge monitoring procedure to reading comprehension, and seemed to have little effect on studies of college learning or performance estimation. The findings indicated that, with the possible exception of relationships with reading comprehension, use of the word list and vocabulary alone appear to be effective in estimating metacognitive knowledge monitoring, whether the text passage is used or not.
The distinction between explicitly and implicitly defined words was expected to be useful only in those studies in which students read the text passage. The results of those investigations indicated that there were few differences between these types of words. Because neither the use of the text passage nor the distinction between the two types of words seemed to affect the results, it seems prudent to abandon that distinction in future research.

Implications for Training Research and Instruction

The results indicated that use of the text passage did not add much variance to the use of the KMA as an appraisal instrument. It may, nevertheless, be interesting to use the passage in future research to study the applicability of the KMA for research on the training of knowledge monitoring. If the word list and vocabulary test are used as pre-post measures, the text passage could be interspersed to help students learn the meanings of those words about which they had made incorrect knowledge estimates. Different levels of instructional support (Tobias, 1989) could be used to help students learn the meanings of the words they had judged incorrectly.

Use of the text passage makes it possible to implement a training strategy featuring maximal prompting in the form of very active instructional interventions at the beginning and fading those out until the passage alone is presented without any prompts. The interventions could include such procedures as: urging students to provide definitions or synonyms for the words, asking them to rephrase the clauses containing the target words, asking questions about the words, and cueing students that the target words are especially important or that they should pay special attention to them. Of course, research would have to determine whether the suggested interventions actually constitute a hierarchy ranging from maximal to minimal support. It should also be noted that a number of passages, with associated word lists and vocabulary sets, may be needed to develop an effective knowledge monitoring training procedure. Once research has determined the usefulness of the procedures outlined above, they could become an important resource to help teachers at all levels improve the knowledge monitoring of their students.

In addition to the possible usefulness of the instructional interventions described above for training, they could enhance the similarity of the KMA to dynamic types of assessment, and to students' school learning. Research could then determine whether such interventions improve the knowledge monitoring procedure's relation-
ship to school learning. It should be noted that, giving students an opportunity for new learning before administering or re-administering, the knowledge estimating procedure is likely to be more complex in mathematics or science than it is for vocabulary. Dynamic assessment in these fields would probably require very active instructional interventions before students can improve their knowledge, because few people can master new material in science or mathematics merely by being asked to read a passage and by twice working on problems in that field, or even by the type of interventions suggested above.

Optimal Indicators of the Latent Knowledge Monitoring Construct

Metacognitive knowledge monitoring is a latent construct inferred from the various scores generated by the procedure. Many of the preceding studies combined the + + and - - scores to form a measure of knowledge monitoring accuracy. The combined score seemed to have face validity as the most direct and most theoretically interesting index of knowledge monitoring accuracy. Furthermore, by including the - - scores the combined total seemed independent of students' actual knowledge, because the combined estimate included items answered incorrectly. Scores based on the signal detection paradigm were used in Study II, but seemed to add little to the combination of + + and - - scores used in the other studies. However, the findings of some of the investigations, especially Studies VII and VIII, suggested that differences between groups were obscured when the sub-scores for different categories (+ +, + -, -+, and - - for words defined explicitly or implicitly) were combined.

Ideally, the optimal knowledge monitoring score should be determined empirically, rather than on the basis of its face validity. The four subscores, or eight if the explicit-implicit distinction is used, generated by the procedure should be submitted to procedures such as the analysis of covariance matrices in order to determine which score(s) are optimal indicators of the latent knowledge monitoring construct. Further research is clearly needed with larger samples (perhaps 200-300 students) than previously employed in order to obtain some stability for the results. The data should then be analyzed with LISREL or comparable procedures in order to identify empirically the optimal score of the latent knowledge monitoring construct.

Knowledge and Estimates of Knowledge

Research has indicated that vocabulary scores are one of the most powerful predictors of school learning (Breland, Jones, & Jenkins,
1994; Just & Carpenter, 1987). The knowledge monitoring procedure scores combine both students' estimates of what they know and their actual knowledge. Thus, the + + score is a composite of both word knowledge, determined by the raw score on the vocabulary test, and the students' estimates of that knowledge. Each of the studies described above examined whether the estimates contributed independent variance above that accounted for by students' knowledge. Operationally, this question was analyzed by comparing the variance accounted for by correct estimates (+ + and - - combined) and those representing only the number correct on the vocabulary test (+ + added to - +). Table 7 summarizes these results for each of the studies.

Table 7 indicates that in Studies V, VII, and VIII (four comparisons) knowledge alone, determined by raw score on the vocabulary test, accounted for more variance (ranging from 1-17%) than the estimates. Also, there seemed to be little difference between actual knowledge and estimates in Study III. When knowledge estimates of college students taking introductory psychology classes were related to their Psychology AP scores, the effect size for knowledge alone was 13% (Study VII) and 17% greater (Study VIII) than for knowledge estimates. When relationships between indices of introductory psychology students' in-class performance and KMA scores were analyzed (Study VIII) the effect size for knowledge alone was 6% greater.

It is not unusual for vocabulary knowledge, even in an unrelated domain, to be an important predictor of students' grades in college exams, such as the multiple-choice test and the AP examination used in Studies VII and VIII. Vocabulary scores in domains not directly related to the curriculum have been powerful predictors of all types of school learning (Breland, Jones, & Jenkins, 1994; Just & Carpenter, 1987), and findings that they were highly related to how much students learned in a psychology course (determined by either the AP exam or in-class tests) were not surprising. Furthermore, because students had little prior experience with the content of the AP examination they had no basis for estimating their performance on that test. Therefore, in such instances it is reasonable that actual knowledge may be more important in determining students' achievement than estimates of that knowledge.

Knowledge estimates accounted for more variance in seven studies, nine comparisons (ranging in effect size or $R^2$ from 1% to 58%, with a median of 4% more variance), compared to knowledge alone. The largest differences occurred in the study of need for feedback where vocabulary raw score accounted for an insignificant 4% of the variance, and accurate knowledge monitoring estimates accounted
Table 7. Summary Comparing Results for Metacognitive Knowledge Estimates and Actual Knowledge.

<table>
<thead>
<tr>
<th>Study</th>
<th>Results Comparing Metacognitive Estimates (KMA) and Number Correct (Raw scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KMA accounted for 4% more variance than raw scores.</td>
</tr>
<tr>
<td>2</td>
<td>KMA accounted for 5% more variance than raw scores.</td>
</tr>
<tr>
<td>3</td>
<td>Correlations similar for KMA and raw scores</td>
</tr>
<tr>
<td>4</td>
<td>Combined GPA differentiated KMA scores, effect size = .07, raw scores NS (effect size = .03). English GPA differentiated KMA scores, effect size = .07, raw scores .06. Humanities GPA differentiated KMA scores, effect size = .09, raw scores .05.</td>
</tr>
<tr>
<td>5</td>
<td>Vocational HS Low and High GPA groups differed on KMA, effect size = .14 and on raw scores, effect size = .16. Predicted, postdicted, and actual final exam score = ns for KMA &amp; raw scores.</td>
</tr>
<tr>
<td>6</td>
<td>Difference between HS Ss and dropouts greater on KMA, effect size = .13 than on raw scores, effect size = .10.</td>
</tr>
<tr>
<td>7</td>
<td>AP data and final grade related to KMA, effect size = .14, and raw scores, effect size = .27.</td>
</tr>
<tr>
<td>8</td>
<td>AP data and final grade data related to KMA, effect size = .16, and raw scores, effect size = .33. Class test data related to KMA, effect size = .17, and raw scores, effect size = .23.</td>
</tr>
<tr>
<td>9</td>
<td>KMA r² with Metropolitan score = .58, raw score = .27.</td>
</tr>
<tr>
<td>10*</td>
<td>Estimates r² with need for feedback = .62, raw score .04 (ns).</td>
</tr>
<tr>
<td>11</td>
<td>Differences between regular, LD, &amp; ADHD Ss greater with KMA than raw scores, effect sizes = .18 compared to .15 for raw scores.</td>
</tr>
</tbody>
</table>

ns = nonsignificant.

* Could not be determined.
for a substantial 62% of the variance! Of course, that finding should be replicated on larger samples. Nevertheless, it seems reasonable that need for feedback should rely more heavily on students' estimates than on their knowledge.

Another large difference between the contributions of estimated and actual scores occurred in Study IX, one of the math studies. Estimates of the number of problems that could be solved accounted for 31% more variance than the problems actually solved. The findings of Study IX were replicated substantially in Study X; unfortunately a computer malfunction made it impossible to compare the estimated and actual scores in that investigation. Although the math studies clearly need replication, the findings suggest that knowledge estimates may be more powerful predictors of success in that domain than in vocabulary.

One possible reason for the substantial effects in mathematics compared to vocabulary may deal with domain similarity. That is, knowledge estimates in math were made from content that was highly similar to the types of problems encountered during math instruction. As indicated above in the discussion of the performance expectation studies, the vocabulary words used in the research were not similar to the domains in which instruction occurred, or to other types of external criteria, perhaps leading to somewhat weaker effects. That interpretation is supported by findings from several of the investigations. In Study I, relating the declarative word knowledge and estimates of that knowledge to reading comprehension, the highest relationships were found for KMA scores after students had read the text passage in which the vocabulary words were defined. That procedure was obviously very similar to the task students face in reading comprehension tests. Furthermore, in Study III social science and science had the lowest relationships with KMA scores, and the effects for social science and science were insignificant in Study IV. Because the KMA materials were developed to be quite general, they were probably dissimilar to what students learned in these more technical areas. These results suggest that the KMA has stronger effects within a domain, rather than across domains. Schraw, Dunkle, Bendixen, and Roedel (1995) found that knowledge monitoring had both domain specific and domain general attributes. Further research is needed to clarify the domain specific and/or domain general characteristics of the KMA.

Another possibility accounting for the more positive results in the studies involving mathematics deals with the perceived difficulty of the subject. Everson, Tobias, Hartman, and Gourgey (1993) found
that students perceive mathematics to be the second most difficult subject, right after science. Conceivably, as suggested below, students’ estimates of their knowledge in more difficult domains are less automatic and involve more reflection about their prior experiences than in simpler areas. Students’ confidence and/or their anxiety about these fields may also affect their estimates. Further research is clearly needed using materials drawn from mathematics, science, and other fields to study both this question and the issue of domain generality-specificity.

**Difficulty and Knowledge Monitoring Procedure**

Little information about the difficulty of the various vocabulary and mathematical materials was available prior to their use in any of the studies. This may well have contributed to some of the variable results. It seems reasonable that estimates of knowledge based on students’ thoughtful consideration of what they know and do not know would be more substantially related to other variables than estimates made more or less automatically. Rapid answers made with little reflection are most likely when students respond to materials that are very easy for them. Wrong estimates for such relatively automatic responses probably indicate careless errors, rather than failures of well-considered estimates. More difficult materials may also evoke nonreflective responses, because students may feel that they neither know nor care about what the correct answers to such questions are. Items of moderate difficulty, about which students may have partial knowledge that can be extended by exerting some effort, would appear to be most likely to elicit well-considered responses reflective of students’ metacognitive knowledge monitoring ability.

Item difficulty is also of importance in considering the different KMA scores. Of the four scores generated by the procedure, the greatest number of responses fell into the ++ category in the studies described above. It may be assumed that more difficult items would yield more -- and +- responses, increasing their reliability and the likelihood that they could contribute more variance to the discrimination between accurate knowledge monitors and their less accurate peers. Furthermore, having more items in the -- category will reduce the similarity between estimates and number correct for two reasons: First, such response represents accurate estimates but no knowledge about the item, and second, more -- items leaves a smaller percentage of ++ items.

In future research these expectations about the effects of varying item difficulty levels should be tested by using items with a previ-
ously determined range of known difficulty. It could be hypothesized that the most useful metacognitive knowledge estimates are likely to be generated from materials of moderate difficulty, and that more difficult items will increase the distinction between KMA accuracy and number correct on any of these procedures.

Relationship to Metamemory Research

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

FOK research was originated by Hart (1965) who asked general information questions of students who, after failing to recall an item, had to make a judgment regarding their FOK about that item. Finally, they were asked to select an answer from a set of distractors. The procedure has been extended to asking students to guess if they could recall words learned in a paired associate task (Hart, 1967; Ryan, Petty, & Wentzlafl, 1982). Nelson, Gerler, and Narens (1984) also extended the FOK research to students’ ability to relearn, and to perceptual identification tasks. Reder and Ritter (1992) investigated whether students opted either to retrieve or calculate mathematical problems, and the latency and accuracy of these processes. A review of FOK research indicated that “a large number of studies confirmed that (students).... unable to retrieve a solicited item from memory can estimate with above chance success whether they will be able to recall it in the future, produce it in response to clues, or identify it among distractors....The standard finding is that the predictive validity of FOK judgments is above chance, though far from perfect” (Koriat, 1993, p. 609-610).

The FOK and JOL paradigms differ from the present research in a number of ways. First, the FOK judgments are typically required after a recall failure, rather than after every stimulus presentation. Second, in FOK or JOL research no attempts are usually made to enable students to learn and/or correct their knowledge of the stimuli, as they were in some of the present research. Third, the purposes of
the metamemory research are to clarify the mechanisms accounting for FOK and JOL, rather than to use the scores as a measure of metacognitive knowledge monitoring to be related to different variables of importance in students' school learning.

Suggestions for Further Research

A number of recommendations for further research have been made earlier; additional suggestions that do not pertain directly to the previous discussion are made here. The positive findings relating knowledge monitoring to need for feedback suggest that studies of similar variables relating the procedure to processes of importance in school learning may be fruitful. For example, forgetting what has been learned in school may be related to knowledge monitoring. It could be inferred that students with good knowledge monitoring abilities, by having a clear sense of what they know and do not know, may be able to retrieve more prior learning than those who have a less secure grasp of what they know and do not know and, hence, may have greater difficulty retrieving prior learning. A pilot study of the knowledge monitoring-forgetting relationship provided substantial support for that reasoning, and will soon be followed up.

The relationship between knowledge monitoring and the effect of distractibility is another fruitful area for investigation. Even though there is a great deal of anecdotal evidence that students are readily distracted from their studies, it has been surprisingly difficult to divert students in investigations specifically designed for that purpose (Slater, 1968; Tobias, 1973). Although some of that variability may be attributable to motivational phenomena (i.e., the interest level of both the primary and distracting materials seems to be important in determining whether students are successfully diverted from their studying; Tobias, 1973), students' knowledge monitoring abilities may also help to determine whether students are distracted. Students with an accurate grasp of their knowledge should find distractions less disruptive from their work than those with a hazier notion of what they know and do not know.

Research should also be conducted relating knowledge monitoring to depth of processing (Craik & Lockhart, 1972). Students should be able to distinguish between the known and unknown more accurately if the learning was processed at a deep, rather than shallow level. Deeper processing should enhance students' knowledge monitoring ability, and it could be predicted that students will make more accurate distinctions between the known and unknown on material
they are induced to process deeply, either by experimental manipulations or instructions, rather than at a shallow level.

Learning in complex domains, such as science, engineering, or making diagnoses in medicine or other fields, often requires that students bring substantial amounts of prior learning to bear in order to understand and acquire new knowledge, and/or solve problems. Some prior learning may be recalled imperfectly, or may never have been completely mastered during initial acquisition. Students who can accurately distinguish between what they know and do not know should be at an advantage while working in such domains, because they are likely to review and try to relearn imperfectly mastered materials needed for particular tasks more readily than students who are less accurate in making such differentiations.

REFERENCES


4. METACOGNITIVE KNOWLEDGE MONITORING


