Instructional Science **26:** 65–79, 1998.

The ability to estimate knowledge and performance in college: A metacognitive analysis*

HOWARD T. EVERSON[1] & SIGMUND TOBIAS[2]
[1]*The College Board, New York;* [2]*City College of New York, New York, U.S.A.*

Abstract. While in college students learn a great deal of new knowledge, and over time successful students learn to update their knowledge as new concepts, facts, and procedures are acquired. The metacognitive ability to accurately estimate one's knowledge was hypothesized to be related to academic achievement in college. The two studies reported in this paper examined the relationship between a measure of metacognitive word knowledge (the KMA) and performance in college. Using undergraduate gpa in a number of academic domains as criterion measures, this research provides support for the validity of the KMA as a predictor of success in college. Suggestions for further research relating performance on the KMA to learning in complex domains are offered.

Metacognition has been defined as the ability to monitor, evaluate, and make plans for one's learning (Flavell, 1979; Brown 1980). The literature in this area identifies two distinct aspects of metacognition, knowledge about cognition and the regulation of cognition, with both viewed as important for effective learning (Brown, 1987; Garner & Alexander, 1989; Jacobs & Paris, 1987). Indeed, researchers have reported differences in metacognitive abilities between capable and less capable learners (see, for example, Baker, 1989; Brown & Campione, 1986; Garner & Alexander, 1989; Pressley & Ghatala, 1990). In general, students with effective metacognitive skills accurately estimate their knowledge in a variety of domains, monitor their on-going learning, update their knowledge, and develop effective plans for new learning. Though widely recognized as important, assessing individual differences in metacognition has proven to be both difficult and time consuming (O'Neil, 1991; Schwartz & Metcalfe, 1994), and remains an obstacle to the advance of research.

Typically, assessments of metacognition rely either on inferences from classroom performance, or ratings based on interviews of students who are

* An earlier version of this paper was presented at a symposium entitled *Issues in Metacognitive Research and Assessment* at the annual meeting of the American Educational Research Association, April 1995, San Francisco, CA, USA.

questioned about their knowledge and cognitive processing strategies, or on analyses of "think-aloud" protocols (Meichenbaum, Burland, Gruson, & Cameron, 1985). Recently, a number of self-report measures of metacognition (Everson, Hartman, Tobias, & Gourgey, 1991; O'Neil, 1991; Pintrich, Smith, Garcia, & McKeachie, 1991; Schraw & Dennison, 1994) have been developed. For the most part, these measures are more efficiently administered and scored than "think aloud" protocols. Unfortunately, the use of self-report measures raises questions of validity (see Schwartz & Metcalfe (1994) for a review of these methodological issues). In light of these concerns, it is not surprising that little research has been conducted on the metacognitive processes related to learning in adults, looking, for example, at those in college or in advanced instructional or training programs, where instructional time less easily accommodates research. Thus, more efficient measures of metacognition are needed not merely to satisfy psychometric standards (although important), but because they would permit research in settings where instructional time is less flexible, such as college classrooms and training courses. In this paper we introduce a method for assessing sudents' knowledge monitoring ability (referred to as the KMA) and relate those scores to their learning and performance in college. Before presenting our results, it may be useful to establish the context for investigating the relationship between metacognition and complex learning in environments such as college and industry-based training courses.

Metacognition and learning

In college students learn a great deal of new knowledge, and are faced, at times, with classroom and laboratory situations that require them to learn material and apply problem solving skills in new and innovative ways. The literature on human metacognition makes a compelling case for its importance in these learning and training environments (Bjork, 1994; Davidson, Deuser, & Sternberg, 1994). Accurate monitoring of new learning enables students with effective metacognitive strategies to concentrate on new content and adjust their learning goals. In college classrooms or advanced training programs, for example, the learner usually has to master a great deal of new knowledge in a limited amount of time. Moreover, learning in classrooms or other structured training environments is often dynamic, with knowledge and information being acquired and updated frequently. Clearly, those who accurately distinguish between what they have already mastered and what is yet to be learned have an advantage in these situations, since they can be more strategic and effective learners. Yet many students have ineffective metacognitive strategies. It is important, therefore, to evaluate students' metacognitive abilities and target instruction to the development of these key learning strategies.

Monitoring knowledge

Given the premise outlined above, we assumed that knowledge monitoring accuracy, an ability presumably involved in the regulation of cognition, would be related to learning in complex environments and reflected in indices such as students' grades in college. Thus, we developed a technique for assessing this metacognitive dimension that conjointly evaluates students' self-reports of their knowledge in a particular domain (e.g., verbal ability) and their performance on an objective measure of knowledge in that domain (see, for example, Tobias et al., 1991; Everson, Smodlaka, & Tobias, 1994; Tobias & Everson, 1996; Tobias & Everson, in press). The basic strategy is to assess knowledge monitoring by evaluating the differences between students' estimates of their knowledge in a particular domain (both procedural and declarative) and their actual knowledge as determined by performance on a test. In the prototypical KMA, students are asked to estimate their knowledge (e.g., in the verbal domain they identify words they know or do not know from a word list, or in mathematics it is problems they expect they can solve) and these estimates are contrasted with their performance on a standardized test containing many of the same words or math problems. Differences between students' estimates and their test performance provide an index of knowledge monitoring ability. This method is similar to methods used in research on metamemory (Nelson & Narens, 1990), reading comprehension (Glenberg, Sanocki, Epstein & Morris, 1987), and psychophysics (Green & Swets, 1966). A brief description of our use of the KMA in an earlier study (Everson et al., 1994) serves as an illustration.

In an effort to understand better the relationship between metacognition and reading comprehension, the KMA was administered to 169 college students. Each was given a list of 33 words and asked to indicate the words they knew and did not know. This was followed by a vocabulary test based on the same words. The KMA generated four scores, including estimates that the word was: a) known and correctly identified on a subsequent vocabulary test [++]; b) known, yet incorrectly identified on the test [+−]; c) unknown, yet correctly identified on the test [−+]; and d) unknown and incorrectly identified on the test [−−]. Within this framework the [++] and the [−−] scores represented accurate metacognitive estimates of vocabulary word knowledge, while the two other measures [+−, and −+] represented inaccurate knowledge monitoring estimates. The results indicated that college students' accurate metacognitive judgments, both the ++ and −− scores, were positively correlated with their scores on a standardized measure of reading comprehension (i.e., the Descriptive Test of Language Skills, 1979), $r = .46$ and $-.43$, respectively. Encouraged by these findings, we adapted the KMA for use in an extensive program of research (Tobias & Everson, 1996; Tobias & Everson, in press).

In light of the importance of verbal reasoning and general word knowledge in a variety of college courses such as English, humanities, and the social and behavioral sciences, we expected KMA scores in the verbal domain to be related to grades in those courses. Thus, the purpose of the two studies reported below was to examine the empirical relationships between and among the KMA scores and indices of learning in college, i.e., grade point averages in various courses, and to investigate whether KMA scores would at all predictive of achievement in courses requiring general verbal abilities.

Study I

Participants

The sample consisted of 139 students attending a large urban university, though only 84 participants completed all the materials during two sessions. A portion of the sample (11%) were students seeking a college degree in nursing. The nursing students ($N = 47$; $N = 33$ with complete data) were recruited from a class serving as the orientation course in a nursing program. The remainder ($N = 92$; $N = 51$ with complete data) was recruited from a freshman orientation course.

Materials

A revised version of the KMA developed in prior research (Everson et al., 1994) was used in this study. In addition to minor editorial revisions of the expository text, a narrative version of the passage was also developed in order to examine the effect of situational interest on metacognition. A total of 38 words was defined in the revised versions of the text, 19 words were explicitly defined (e.g., “Coronary or heart disease . . . ”), and another 19 received implicit definitions (e.g., “Epidemiologists who have compared the prevalence of heart disease in the United States and in other countries . . . ”). Explicit or implicit definitions were determined by two independent judges who rated all words. When there was any disagreement about a particular word, the judges conferred and the passage was modified to eliminate the disagreement. The word list and vocabulary task were also modified to contain an equal number of explicitly and implicitly defined words. A multiple choice vocabulary test was developed, containing the correct choice and three distractors for the 38 items on the word list.

Procedures

The KMA word list and vocabulary test, coefficient alpha = .80 (Cronbach, 1951), were administered first in a group session. The two versions of the text were then randomly assigned to students in a second experimental session, followed by a re-administration of the KMA word list and vocabulary test. In each instance, the experimental materials were administered during class in the presence of the instructors.

Results and discussion

The correlations between total score on both administrations of the vocabulary test, based on the 84 students who completed the test on both administrations, was r = .75. (This correlation, however, should not be interpreted as an estimate of test-retest reliability (Crocker & Algina, 1986), because students read the text passage, from which the meaning of the words could be inferred, immediately before the second administration of the vocabulary test.) Students' estimated word knowledge and performance on the vocabulary test were determined for both administrations. Two scores were computed for each administration: the total number of correct (words in the [++] and [−−] categories) and incorrect estimates (words in the [+−] and the [−+] categories). Preliminary analyses revealed no differences between students assigned to the expository or narrative text versions, or between ex- and implicitly defined words. Therefore the data for both text versions and both types of words were pooled in subsequent analyses. For this sample of 84 students, 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 words once they read the passage and updated their word knowledge. 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 1.

Since 92 participants were freshmen in their initial 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 1 also shows the correlations for metacognitive

Table 1. Correlations between knowledge monitoring scores, raw scores, and overall grade-point average in different subject areas.

Variables	*Administration 1*			*Administration 2*		
Group	*Correct Estimate*		*Raw Score*	*Correct Estimate*		*Raw Score*
Total GPA	*N*	*r*	*r*	*N*	*r*	*r*
Total	101	.20*	.01	94	.09	–.00
Freshmen	65	.09	–.25	61	–.10	–.21
Nurses	36	.28*	–.37*	33	.19	.17
English GPA						
Total	72	.30**	.19	63	.19	.05
Freshmen	53	.31**	.10	48	.00	.16
Nurses	19	.25	.33	19	.45*	.44*
Humanities GPA						
Total	82	.26**	.04	74	.13	.00
Freshmen	52	.12	–.21	46	–.11	.22
Nurses	30	.48**	.40*	28	.35*	.24
Science GPA						
Total	65	.18	–.01	63	.03	–.07
Freshmen	28	.11	–.30	27	–.28	–.47
Nurses	37	.26	–.42*	26	.18	.26
Social Science GPA						
Total	64	.18	.26	63	.24	–.26*
Freshmen	26	.15	.10	29	.14	.18
Nurses	38	.09	–.31	34	.14	.10

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

estimates and vocabulary raw scores, (i.e., the number correct on the vocabulary test), separately. Finally, the different number of cases in the various cells of Table 1 should also be noted. The correlations shown in Table 1 are generally positive and frequently significant, ranging in magnitude from low to moderate. The results support the concurrent validity of the KMA with respect to its relationship to learning in college. As expected, relationships between knowledge monitoring scores and GPA in English were generally highest, followed by humanities courses and the combined GPA, while correlations with social science and science GPAs were generally lower. The largely non-significant relationships with social science courses were surprising, since it had been assumed that these courses usually contained material and vocabulary which was less technical or unfamiliar to students than natural science courses. Perhaps grades in these courses, like those in the natural sciences, reflected greater domain-specific knowledge than found in the English and humanities courses.

The significance of the correlations reported in Table 1 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 either administration of the materials, leading to variability in the predictors. Second, it is well known that college grades are often unreliable (Werts, Linn, & Jöreskög, 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 1; b) when similar courses were taken they were taught by different instructors; and c) the differences in students' major fields of study. As expected, the correlations between KMA scores and grades in English were generally higher, and more frequently significant, than those of any other subject. The findings indicate that the metacognitive knowledge monitoring scores are related to students' ability to learn materials from somewhat different domains than the ones on which the KMA was based.

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 from the first. The findings did not support these expectations. While the increase in vocabulary score after reading the text was statistically significant, it indicated that 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.

The results shown in Table 1 also indicate that the correlations with number correct on the vocabulary scores were generally similar to the relationships with correct knowledge monitoring estimates. Due to the varying Ns in the different cells, the differences in the correlations were examined with a *t* test developed by Hotelling (Tukey, 1977). For the correlations with GPA based on the both administrations, using the total group, the KMA scores were higher seven times (one difference was significant at $p < .05$), while the correlations based on raw scores were higher three times (none significantly so). For freshmen, the correlations with the KMA scores were higher twice, but not significantly so, while correlations with raw vocabulary score were higher eight times (two were significant at $p < .05$) Finally, for nursing students, correlations with scores were higher five times (none significant), while relationships based on raw vocabulary scores were higher five times (one significant $p < .05$). In contrast to the findings from earlier studies (Tobias et al., 1991; Tobias & Everson, 1996; Tobias & Everson, in press), the KMA scores appeared to add little independent variance to the relationship with grades beyond that accounted for by the number correct on the vocabulary test.

Study II

The preceding study was concerned with the issue of concurrent validity and examined the correlations of KMA with students' prior learning in college. To extend this work we undertook a second study that investigated the KMA's predictive validity, by examining whether the metacognitive estimates would predict how well entering students would perform academically during their first year of college.

Procedures and participants

The materials used in Study II were identical to those used in Study I. They were administered while students attended a pre-freshman skills program before beginning their first semester of college. Achievement was determined by obtaining students' grade point averages (GPAs) at the end of their first year of college in the same subjects examined in the prior study: English, humanities, sciences, and social sciences, as well as the combined GPA. The sample consisted of 115 students (59 female) participating in a skills program intended for students considered at risk of doing poorly in their first year of college.

Each participant completed all of the study materials and took similar types of courses. High- and low-achievement groups were created by dividing students at the GPA median for the different academic areas and for the combined GPA. Then differences in knowledge monitoring ability between the groups were examined. Mixed between- and within-subjects analyses of variance were computed to determine the significance of the differences in estimates of word knowledge between groups above and below the GPA median. At the conclusion of the freshman year, it was determined that 95 of the 115 original participants had completed some courses at the college.

Results and discussion

The number of correct estimates students made of their word knowledge was determined. As in the prior studies, correct estimates were defined, by combining the [++] and [−−] categories. Preliminary analyses again indicated that there were no differences between the results obtained for the expository and narrative passages, nor between the words defined explicitly or implicitly. Therefore, these data were pooled for the succeeding analyses.

ANOVA indicated that, as expected, students above the median GPA (N = 48) made significantly more accurate overall estimates of their knowledge (M_H = 49.2), $F(1, 93) = 6.42$, $p < .05$, on both administrations than those below the median (N = 47, M_L = 45.8,); the size of that effect, determined by eta^2 (η^2; Kennedy, 1970) was .065. Also as expected, there was a significant difference between the mean of the first administration (M_1 = 22.9) and the mean of the second administration (M_2 = 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 mean differences between the high (M_H = 43.2) and low GPA groups (M_L = 39.3) on the vocabulary test were not significant, $F(1, 93) = 2.73$, $\eta^2 = .029$, while the differences between the means of the first (M_1 = 17.7) and second administrations (M_2 = 24.5) were highly significant $F(1, 93) = 198.04$, $p < .001$, $\eta^2 = .68$. Again, there were no interactions between the variables.

High- and low-achieving groups in English, humanities, science, and social science courses were also identified by dividing the students at the median GPA for each of these subject areas and examining the significance of differences on the number of correct estimates of word knowledge. In English, the overall differences in the accuracy of the estimates between students above (M_{EH} = 48.9) and below the median (M_{EL} = 45.4) were significant, $F(1, 82) = 6.18$, $p = .02$; $\eta^2 = .07$, as were the differences between the first (M_{E1} = 45.6) and second administrations (M_{E2} = 48.7), $F(1, 82) = 11.92$; $p < .01$; $\eta^2 = .127$). Furthermore, there was an interaction between groups and administra-

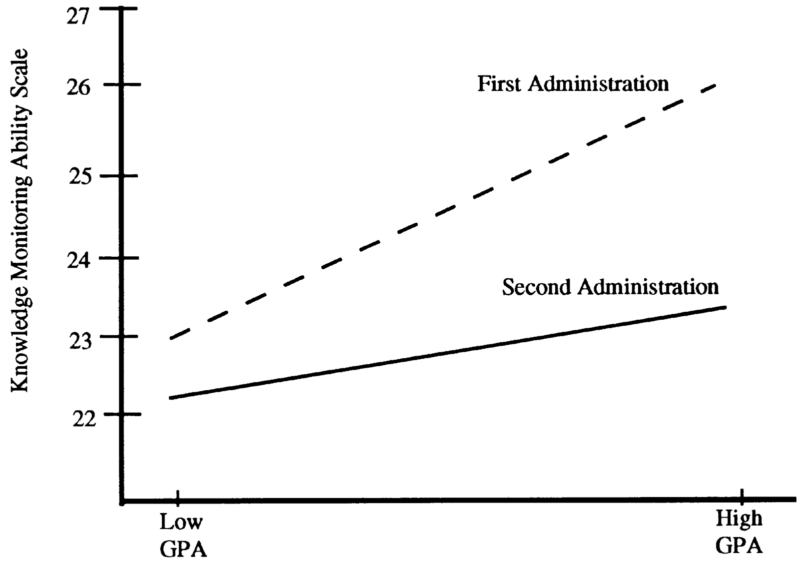

Figure 1. Interaction of English GPA groups, correct KMA estimates, and test administrations.

tions, $F(1, 82) = 4.41$, $p < .05$; $\eta^2 = .051$. The interaction, as shown in Figure 1, suggests that while the accuracy of both groups' estimates of known and unknown words increased from the first to the second administration, higher achieving students made greater gains. A similar analysis was computed for number correct on both vocabulary test administrations. The finding indicated that the difference between the high- ($M_{EH} = 42.9$) and low-GPA groups ($M_{EL} = 38.9$), $F(1, 82) = 5.43$; $\eta^2 = .062$, was slightly smaller than that determined when the KMA scores were used, but there was a stronger effect for differences between the first ($M_{E1} = 18.0$) and second administrations ($M_{E2} = 23.6$), $F(1, 82) = 169$, $p < .001$; $\eta^2 = .673$; there was no evidence for interaction in these results.

Similar analyses were made for students above and below the median in humanities courses (Art, History, Music, Philosophy, World Civilization, World Humanities, and World Arts). Differences between the means of the high ($M_{HH} = 49.4$) and low humanities GPA groups ($M_{HL} = 45.3$) were also significant, $F(1, 81) = 7.96$, $p < .01$; $\eta^2 = .089$, as were the differences between first ($M_{H1} = 23.0$) and second administrations ($M_{H2} = 24.5$), $F(1, 81) = 9.94$, $p < .001$; $\eta^2 = .109$; there was no interaction. The same type of analysis was also computed for number correct on the first and second vocabulary tests; again it revealed somewhat smaller differences between the high- ($M_{HH} = 43.1$) and low GPA groups ($M_{HL} = 39.0$), $F(1, 81) = 4.18$, $p < .05$; $\eta^2 = .049$) and larger differences between the first ($M_{H1} = 17.8$) and second administration ($M_{H2} = 23.4$), $F(1, 81) = 179.2$, $p < .001$; $\eta^2 = .689$, than the results for knowledge monitoring scores. There were no significant differences between the science or the social science GPA groups using either the knowledge monitoring scores or the vocabulary raw scores.

The relationships between metacognitive scores and GPA were generally similar to those reported in Study I, supporting the predictive validity of the KMA scores. In contrast with Study I, in which both the KMA scores and the vocabulary raw scores had fairly similar patterns of relationship, the metacognitive scores had a significant effect on overall GPA, whereas the raw scores did not. Furthermore, the KMA scores accounted for more variance between groups than did the number correct on the vocabulary test in two of three other comparisons, supporting the construct validity of the procedure.

General discussion

The findings of the two studies summarized above provide support for the validity of the KMA, both in terms of the construct of metacognitive knowledge monitoring and the predictive validity of the assessment procedure. The results suggest that the procedure has some generality across different samples of college students, various types of content, as well as different types of vocabulary. In general, the KMA scores seemed to more successfully differentiate the capable students, whose grades were above the median, from those less able than did the raw vocabulary scores, replicating the findings reported elsewhere (Tobias & Everson, 1996). The knowledge monitoring scores accounted for anywhere from 1 to 4 percent more variance than did similar analyses using the raw vocabulary scores.

It was also interesting that the analysis of differences in the raw vocabulary scores 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 KMA in that most students learned some new words from reading the passage, though their knowledge monitoring ability was not equally enhanced. However, it should be noted that the results for the English grades indicated that there were greater increases in knowledge monitoring ability for capable students than for their less able peers (see Figure 1). These findings suggest that while all students increased both their demonstrated knowledge of vocabulary and their knowledge monitoring ability from the first to second administration, the increases monitoring ability were greater for more capable students (i.e., those whose English grades were above the median). Apparently there was a greater degree of improvement in such students' metacognitive skills than in those of their less able colleagues.

Research has indicated that vocabulary test scores are one of the most powerful predictors of school learning (Breland, Jones, & Jenkins, 1994; Just & Carpenter, 1987). KMA scores combine both students' estimates of what they know and their actual knowledge. Thus, the ++ score is a composite of both actual word knowledge, determined by the raw score on the vocabulary test, and the students' correct estimates of that knowledge. Each of the studies described above examined whether the KMA estimates contributed independent information beyond that accounted for by students' actual word knowledge. Operationally, this question was analyzed by comparing the variance accounted for by correct estimates (++ and −− combined) with the variance accounted for using only the number correct on the vocabulary test (++ and −+). In general, the effect size for actual word knowledge alone was greater (about 13%) than for the knowledge estimates. It is not unusual for knowledge of vocabulary, even in unrelated academic domains, to be an important predictor of students' grades in college. Vocabulary scores based on words not directly related to a particular course or curriculum have been shown to be powerful predictors of all types of classroom learning (Breland et al., 1994; Just & Carpenter, 1987). Further research is obviously needed to clarify these effects.

Learning in complex domains such as science and 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 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, since they are more likely to review and try to relearn imperfectly mastered materials needed for particular tasks, compared with those who are less accurate in estimating their own knowledge. In view of the fact that the existing knowledge monitoring scores accounted for little variance in science grades, it would be useful to develop a KMA procedure in the science domain to determine its relationship to achievement in science and engineering.

Further research is also needed to determine the relationships between the KMA procedure and self-report measures of metacognition, study skills, and self-regulated learning. These constructs have some similarity to the KMA procedure and positive relationships should be obtained. Finally, the relationship between knowledge monitoring ability and measures of intelligence should be investigated. Sternberg (1991) has suggested that metacognition should be a component of intelligence tests; presumably those who consider metacognition an executive process (Borkowski, Chan, & Muthukrishna, in press) would also agree with that recommendation. Research findings

(Schraw, in press) indicate that academically able students have higher knowledge monitoring ability than those less able. Therefore, positive relationships between the KMA procedure and measures of general intellectual ability may be expected.

Several factors are likely to have reduced the magnitude of the effects and the generalizability of the results to other groups of college students. As in the first study, the participants in the pre-freshmen program were considered to be at risk for poor performance in college. This may have reduced the range of achievement for the sample and, therefore, may also have reduced the differences in the knowledge monitoring ability between the groups. Furthermore, even though data were *not* collected in classes of the pre-freshmen skills program devoted exclusively to English as a Second Language (ESL), some of the students were enrolled in both ESL and other courses, and thus ended up as part of the sample. The presence of non-native English speakers could also have reduced group differences in this study. Further research limited to native English speakers may reduce the variability among participants and narrow the group differences.

It should be noted that many of the students in this sample took less than a full-time schedule of courses. That fact 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 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 KMA procedure for a large number of full-time students.

Acknowledgements

Preparation of this paper was partially supported by The College Board. Portions of this research were conducted while the second author was a Visiting Faculty Fellow at the Navy Personnel Research and Development Center in San Diego, in a program sponsored by the American Association for Engineering Education, and the U.S. Navy.

References

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

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A.P. Shimamura, eds., *Metacognition: Knowing about knowing* (pp. 185–206). Cambridge, MA: MIT Press.

Borkowski, J.G., Chan, L.K.S. & Muthukrishna, N. (in press). A process-oriented model of metacognition and executive functioning. In G. Schraw, ed., *Issues in the measurement of metacognition*. Lincoln, NE: Buros Institute / The University of Nebraska Press.

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

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

Brown, A.L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. Weinart & R. Kluwe, eds., *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Erlbaum Associates.

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

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart and Winston.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297–334.

Davidson, J.E., Deuser, R. & Sternberg, R.J. (1994). The role of metacognition in problem solving. In J. Metcalfe & A.P. Shimamura, eds., *Metacognition: Knowing about knowing* (pp. 207–226). Cambridge, MA: MIT Press.

Descriptive Test of Language Skills. (1979). Princeton, NJ: The College Entrance Examination Board.

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

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

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

Garner, R. & Alexander, P. (1989). Metacognition: Answered and unanswered questions. *Educational Psychologist* 24: 143–158.

Glenberg, A.M., Sanocki, T., Epstein, W. & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General* 116(2): 119–136.

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

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

Just, M.A. & Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Needham Heights, MA: Allyn & Bacon.

Kennedy, J.J. (1970). The eta coefficient in complex ANOVA designs. *Educational and Psychological Measurement* 30: 885–889.

Meichenbaum, D., Burland, S., Gruson, L. & Cameron, R. (1985). Metacognitive assessment. In S.R. Yussen, ed., *The growth of reflection in children* (pp. 3–27). NY: Academic Press.

Nelson, T.O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower, ed., *The psychology of learning and motivation* (Vol. 26) NY: Academic Press.

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

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

Pressley, M. & Ghatala, E.S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist* 25: 19–33.

Schraw, G. (in press). *Issues in the measurement of metacognition*. Lincoln, NE: Buros Institute/The University of Nebraska Press.

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

Schwartz, B.L. & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A.P. Shimamura, eds., *Metacognition: Knowing about knowing* (pp. 93–114). Cambridge, MA: MIT Press.

Sternberg, R.J. (1991). Toward better intelligence tests. In M.C. Wittrock & E.L. Baker, eds., *Testing and cognition* (pp. 31–39). Englewood Cliffs, NJ: Prentice-Hall.

Tobias, S. & Everson, H.T. (in press). Assessing metacognitive word knowledge. In G. Schraw, ed., *Issues in the measurement of metacognition*. Lincoln, NE: Buros Institute/University of Nebraska Press.

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

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

Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

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

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