DOCUMENT RESUME

ED 404 373 TM 026 459

AUTHOR O'Neil, Harold F., Jr.; And Others
TITLE Experimental Studies on Motivation and NAEP Test Performance. Final Report. NAEP TRP Task 3a: Experimental Motivation
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE Dec 92
CONTRACT RS90159001
NOTE 222p.
PUB TYPE Information Analyses (070) -- Statistical Data (110)

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS Achievement; *Experiments; Grade 8; Grade 12; *Incentives; *Mathematics Tests; Performance Factors; Pilot Projects; Program Descriptions; Research Methodology; Secondary Education; *Student Motivation; Tables (Data); Test Construction; *Testing; Test Use
IDENTIFIERS *Low Stakes Tests; *National Assessment of Educational Progress

ABSTRACT

The Cognitive Science Laboratory of the University of Southern California has conducted a series of studies on the experimental effects of motivation on a low-stakes (to the student) standardized test. This report summarizes these studies and their results. The test in question is the National Assessment of Educational Progress (NAEP). A series of studies in 1992 investigated the effects of various motivational conditions on the performance of 8th and 12th graders on a subset of items from the NAEP 1990 mathematics test. Several pilot studies were conducted first to select the motivational conditions that might influence performance. The main study compared the effects of financial reward, competition, personal accomplishment, and standard NAEP test instructions on mathematics performance. Results indicate that financial reward can improve the performance of eighth graders. In the 12th grade, no differences were observed among the conditions. The eighth grade findings indicate that test developers may be underestimating the achievement of students when scores on low stakes tests are used as the indicators of achievement. Five appendixes discuss study methodology, instructions, and detailed results. (Contains 93 tables, 103 appendix tables, and 65 references.) (SLD)

C·R·E·S·S·T

National Center for Research on Evaluation, Standards, and Student Testing

National Center for Research on Evaluation, Standards, and Student Testing

Final Deliverable – December 1992

NAEP TRP Task 3a: Experimental Motivation Study

Final Report of Experimental Studies on Motivation and NAEP Test Performance

▸ UCLA Center for the Study of Evaluation

in collaboration with:

▸ University of Colorado

▸ NORC, University of Chicago

▸ LRDC, University of Pittsburgh

▸ The RAND Corporation

National Center for Research on
Evaluation, Standards, and Student Testing

Final Deliverable – December 1992

NAEP TRP Task 3a: Experimental Motivation Study

Final Report of Experimental Studies on Motivation and NAEP Test Performance

Study Director: Harold F. O'Neil, Jr.
University of Southern California/CRESST

U.S. Department of Education
National Center for Education Statistics
Grant RS90159001

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

The work reported herein was supported in part under the National Center for Education Statistics Contract No. RS90159001 as administered by the U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the National Center for Education Statistics or the U.S. Department of Education.

Acknowledgments

We would like to thank all of the students who participated in these studies, and the school and district personnel who facilitated access to those students. We owe special gratitude to Steve Rhine (UCLA, Graduate School of Education) and the 14 retired school personnel listed below who administered the tests and questionnaires and distributed feedback and money to the students. Their enthusiasm, energy and competence was remarkable.

Jim Burk	Phyllis Marquart
Elma Dinkins	Roy Nakawatase
Barbara Espinoza	Mary O'Neill
Evelyn Friedman	Lorna Round
Rosalyn Heyman	Cheryl Sparti
Vivian Johnson	Willa Waters
Ranyer Mann	Paul Yokota

Our gratitude also to Josie Bain, CRESST, for her invaluable contribution.

TABLE OF CONTENTS

FINAL REPORT OF EXPERIMENTAL STUDIES ON MOTIVATION AND NAEP TEST PERFORMANCE

EXECUTIVE SUMMARY

Harold F. O'Neil, Jr., CRESST/University of Southern California
Brenda Sugrue, CRESST/University of California, Los Angeles
Jamal Abedi, CRESST/University of California, Los Angeles
Eva L. Baker, CRESST/University of California, Los Angeles
Shari Golan, CRESST/University of California, Los Angeles

Introduction

The Cognitive Science Laboratory of the University of Southern California has a subcontract with the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles to assist in the research on the experimental effects of motivation on the National Assessment of Educational Progress (NAEP). The University of Colorado/CRESST has conducted a study on embedded NAEP tests in a state assessment. In turn, CRESST/UCLA has an existing contract from the National Center for Education Statistics (NCES) to conduct validity studies on NAEP. CRESST/UCLA areas of interest include both assessment and policy issues. The purpose of this report (the Final Report on our USC subcontract) is to document a series of collaborative studies on the experimental effects of motivation on a low-stakes (to the student) standardized test.

The Research Question

One of the major validity questions that has been raised in relation to the National Assessment of Educational Progress (NAEP) concerns the possible impact of motivational factors on the NAEP results. If students are not motivated to perform well on NAEP tests, and if the lack of motivation results in poor performance, then NAEP findings are underestimates of student achievement.

The possibility that NAEP underestimates what students could do if they gave the assessment their best effort has been a concern for some time. Shanker (1990), for example, noted that "one of the most frequently offered

theories about the low NAEP scores is that kids know the tests don't count" and therefore "may decide it's not worth their while to put forth any effort." He went on to argue that because of the importance of NAEP as a source of information about student achievement, "we ought to clear up this question about its validity." Responses to the NAEP mathematics field test questions (Educational Testing Service, 1991) also indicate the need to investigate effort in the context of low-stakes NAEP testing. When asked, "How hard did you try on this test?" 28% of 8th graders responded "Somewhat hard" or "Not at all hard," whereas 51% of 12th graders answered in this manner. Similarly, when asked, "How important was it for you to do well on this test?" 36% of 8th graders responded "Somewhat important" or "Not very important," whereas 62% of 12th graders gave this response.

The Studies

To test the theory that increased motivation to perform well on a NAEP test would be reflected in increased effort and improved performance on the test, a series of studies was conducted in 1992 by UCLA's Center for the Study of Evaluation and its National Center for Research on Evaluation, Standards, and Student Testing (CRESST). The studies investigated the effects of various motivational conditions on the performance of 8th- and 12th-grade students on a subset of released items from the 1990 NAEP mathematics test.

Mathematics was selected because it is a content area that many students not only find difficult but also dislike, want to avoid, or feel anxious about. In addition, mathematics is an area that has been singled out for special attention by its choice as the first content domain in the NAEP Trial State Assessment and for the assessment of the President's and Governors' National Education Goals.

The studies were conducted at two grade levels, 8 and 12. Grade 12 was selected because it is the grade where concerns about motivation are most serious. We did not want to limit the study to that grade, however, because negative effects of low motivation observed at grade 12, if any, might not generalize to other grades. Therefore, we thought it important to replicate the studies at a second grade level. At grade 8, it would be possible to implement some sort of remediation, if desired.

In order to link any observed performance differences to differential investment of effort or to differences in metacognition, anxiety, and perceived ability, these variables were measured via a modified self-assessment questionnaire (O'Neil, Baker, Jacoby, Ni, & Wittrock, 1990) The history of the

development and validation of this instrument is described in detail later in this report.

It was reasoned that the motivational treatments might have different effects on subgroups of students whose performance on NAEP mathematics tests currently differs. Therefore, the studies investigated possible differential effects of the motivational conditions on the performance, and perceived effort, metacognition, mathematics ability, and anxiety of male and female students with different ethnic backgrounds (White, African American, Latino, Asian).

A number of pilot studies were conducted to select the motivational conditions that might influence test performance. (Each of these is described in detail later in the report.) An initial "focus-group study" revealed that both 8th- and 12th-grade students would be motivated by financial rewards to try harder on tests. A second pilot study compared the performance of 8th- and 12th-grade students who received three different financial rewards (or no reward). The study yielded no differences among test scores of 8th- or 12th-grade students who received any of three financial incentives and students who received standard NAEP test instructions. Based on previous research and on our feeling that 50 cents per item might not be enough to motivate Los Angeles teenagers, a financial incentive condition offering a larger reward of $1 per correct item was included in the main study.

A third pilot study investigated the differential effects of various goal orientation conditions. One group of students was told that the goal of the test was to provide a personal challenge and accomplishment (task-oriented goal); a second group was told that the goal was to compare their mathematical ability with that of other students (competitive or ego-oriented goal); a third group was told that the goal of the test was to evaluate the effectiveness of their teachers (teacher-oriented goal); a fourth group in this pilot study got the standard NAEP test instructions. Eighth-grade students (in classes tested first) who were told that the goal was to compare their mathematics ability with that of others obtained higher scores than 8th-grade students who received standard NAEP instructions. However, since this finding was inconsistent with previous research on the relationship of goal orientation and performance (see our literature review), both the personal accomplishment goal and the competitive goal were retained as motivational conditions in the main study.

The main study compared the effects of three experimental motivational conditions (financial reward, competition, personal accomplishment) and standard NAEP test instructions on the mathematics performance of 8th- and 12th-grade students. In addition, for 12th-grade students, a fifth condition was added: Students were offered a certificate of accomplishment if they scored in

the top 10% of their class. The results indicated that the offer of a financial reward can improve the performance of 8th-grade students. The 8th-grade students who were offered a financial reward also reported investing more effort during the test than did 8th-grade students who received the standard NAEP test instructions. Goal orientation manipulations did not result in significant differences on any outcome variable. In 12th grade, no differences were observed in test performance among students who were exposed to the different motivational conditions. However, 12th-grade students who were offered the financial reward reported more metacognitive activity during the test. Treatment did not interact with ethnicity or gender in its effect on any outcome variable in either 8th or 12th grade.

The Implications

The 8th-grade findings indicate that, indeed, we may be underestimating the achievement of students when we use scores on "low-stakes" tests as the indicators of achievement. While offering all students a financial reward for performance on such tests is not practical, there may be other ways of rewarding students for high achievement on such tests that would lead them to invest their maximum effort.

FINAL REPORT OF EXPERIMENTAL STUDIES ON MOTIVATION AND NAEP TEST PERFORMANCE

Harold F. O'Neil, Jr., CRESST/University of Southern California

Brenda Sugrue, CRESST/University of California, Los Angeles

Jamal Abedi, CRESST/University of California, Los Angeles

Eva L. Baker, CRESST/University of California, Los Angeles

Shari Golan, CRESST/University of California, Los Angeles

INTRODUCTION AND SUMMARY

The Cognitive Science Laboratory of the University of Southern California has a subcontract with the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles to assist in the research on the experimental effects of motivation on the National Assessment of Educational Progress (NAEP). The University of Colorado/CRESST has conducted a study on embedded NAEP tests in a state assessment. In turn, CRESST/UCLA has an existing contract from the National Center for Education Statistics (NCES) to conduct validity studies on NAEP. CRESST/UCLA areas of interest include both assessment and policy issues. The purpose of this report (the Final Report on our USC subcontract) is to document a series of collaborative studies on the experimental effects of motivation on a low-stakes (to the student) standardized test.

The Research Question

One of the major validity questions that has been raised in relation to the National Assessment of Educational Progress (NAEP) concerns the possible impact of motivational factors on the NAEP results. If students are not motivated to perform well on NAEP tests, and if the lack of motivation results

in poor performance, then NAEP findings are underestimates of student achievement.

The possibility that NAEP underestimates what students could do if they gave the assessment their best effort has been a concern for some time. Shanker (1990), for example, noted that "one of the most frequently offered theories about the low NAEP scores is that kids know the tests don't count" and therefore "may decide it's not worth their while to put forth any effort." He went on to argue that because of the importance of NAEP as a source of information about student achievement, "we ought to clear up this question about its validity." Responses to the NAEP mathematics field test questions (Educational Testing Service, 1991) also indicate the need to investigate effort in the context of low-stakes NAEP testing. When asked, "How hard did you try on this test?" 28% of 8th graders responded "Somewhat hard" or "Not at all hard," whereas 51% of 12th graders answered in this manner. Similarly, when asked, "How important was it for you to do well on this test?" 36% of 8th graders responded "Somewhat important" or "Not very important," whereas 62% of 12th graders gave this response.

The Studies

To test the theory that increased motivation to perform well on a NAEP test would be reflected in increased effort and improved performance on the test, a series of studies was conducted in 1992 by UCLA's Center for the Study of Evaluation and its National Center for Research on Evaluation, Standards, and Student Testing (CRESST). The studies investigated the effects of various motivational conditions on the performance of 8th- and 12th-grade students on a subset of released items from the 1990 NAEP mathematics test.

Mathematics was selected because it is a content area that many students not only find difficult but also dislike, want to avoid, or feel anxious about. In addition, mathematics is an area that has been singled out for special attention by its choice as the first content domain in the NAEP Trial State Assessment and for the assessment of the President's and Governors' National Education Goals.

The studies were conducted at two grade levels, 8 and 12. Grade 12 was selected because it is the grade where concerns about motivation are most

serious. We did not want to limit the study to that grade, however, because negative effects of low motivation observed at grade 12, if any, might not generalize to other grades. Therefore, we thought it important to replicate the studies at a second grade level. At grade 8, it would be possible to implement some sort of remediation, if desired.

In order to link any observed performance differences to differential investment of effort or to differences in metacognition, anxiety, and perceived ability, these variables were measured via a modified self-assessment questionnaire (O'Neil, Baker, Jacoby, Ni, & Wittrock, 1990) The history of the development and validation of this instrument is described in detail later in this report.

It was reasoned that the motivational treatments might have different effects on subgroups of students whose performance on NAEP mathematics tests currently differs. Therefore, the studies investigated possible differential effects of the motivational conditions on the performance, and perceived effort, metacognition, mathematics ability, and anxiety of male and female students with different ethnic backgrounds (White, African American, Latino, Asian).

A number of pilot studies were conducted to select the motivational conditions that might influence test performance. (Each of these is described in detail later in the report.) An initial "focus-group study" revealed that both 8th- and 12th-grade students would be motivated by financial rewards to try harder on tests. A second pilot study compared the performance of 8th- and 12th-grade students who received three different financial rewards (or no reward). The study yielded no differences among test scores of 8th- or 12th-grade students who received any of three financial incentives and students who received standard NAEP test instructions. Based on previous research and on our feeling that 50 cents per item might not be enough to motivate Los Angeles teenagers, a financial incentive condition offering a larger reward of $1 per correct item was included in the main study.

A third pilot study investigated the differential effects of various goal orientation conditions. One group of students was told that the goal of the test was to provide a personal challenge and accomplishment (task-oriented goal); a second group was told that the goal was to compare their mathematical ability with that of other students (competitive or ego-oriented goal); a third

group was told that the goal of the test was to evaluate the effectiveness of their teachers (teacher-oriented goal); a fourth group in this pilot study got the standard NAEP test instructions. Eighth-grade students who were told that the goal was to compare their mathematics ability with that of others obtained higher scores than 8th-grade students who received standard NAEP instructions. However, since this finding was inconsistent with previous research on the relationship of goal orientation and performance (see our literature review), both the personal accomplishment goal and the competitive goal were retained as motivational conditions in the main study.

The main study compared the effects of three experimental motivational conditions (financial reward, competition, personal accomplishment) and standard NAEP test instructions on the mathematics performance of 8th- and 12th-grade students. In addition, for 12th-grade students, a fifth condition was added: Students were offered a certificate of accomplishment if they scored in the top 10% of their class. The results indicated that the offer of a financial reward can improve the performance of 8th-grade students. The 8th-grade students who were offered a financial reward also reported investing more effort during the test than did 8th-grade students who received the standard NAEP test instructions. Goal orientation manipulations did not result in significant differences on any outcome variable.

In 12th grade, no differences were observed in test performance among students who were exposed to the different motivational conditions. However, students who were offered the financial reward reported more metacognitive activity during the test. In general, treatment did not interact with ethnicity or gender in its effect on any outcome variable in either 8th or 12th grade.

LITERATURE REVIEW

The purpose of this review is to provide a rationale for the set of independent and dependent variables that were selected for investigation in the studies described in this report.

The review is divided into a number of sections. First, the relationship between motivation and achievement is discussed. Second, two educational variables that have been found to influence motivation and performance (rewards and goal orientation) are described. Third, the review provides the

rationale for measuring cognitive processing variables in a study that examines the influence of motivational variables on achievement. Fourth, discussion turns to state test anxiety, a variable operating specifically at the time of test taking and one that affects both cognitive processing and test performance. Fifth, a rationale is developed for examining the differential effects of motivational manipulations on the test performance of different ethnic groups and of male and females. Finally, we discuss the need to report patterns of non-response to test items in addition to performance data.

The Relationship Between Motivation and Achievement

Motivation is a nebulous construct that has been defined as "goal-oriented strivings" (Dweck, 1989) or "the process whereby goal-directed behavior is instigated and sustained" (Schunk, 1990). "Motivation" itself is a latent variable that can only be studied indirectly through variables that seem to give rise to it and that seem to be affected by it. There is a large body of literature on variables that *precede* motivation, such as attributions (Weiner, 1986), expectancies (Eccles, 1983), self-efficacy (Bandura, 1977; Schunk, 1989), perceived control (Stipek & Weisz, 1981), goals (Ames, 1992; Dweck, 1989; Nicholls, 1983), anxiety (Hembree, 1988; Hill & Wigfield, 1984; O'Neil & Abedi, 1992; Wigfield & Eccles, 1989), and variables that *follow* motivation, such as interest (Hidi, 1990), task choice (Kukla, 1978; Nicholls, 1984), effort (Covington & Omelich, 1979; Salomon, 1983), and learning and performance (Helmke, 1989; Uguroglu & Walberg, 1979). However, as d'Ydewalle (1987) has pointed out, "clear-cut results from neat experiments on the impact of motivation on learning [or performance] do not exist" (p. 195).

In the educational context, most existing studies have focused on the influence of characteristics of the classroom learning environment, such as rewards (Deci, 1971, Schunk, 1983), teacher feedback (Brophy, 1981; Butler, 1987; Graham & Weiner, 1986), goal structures (Ames, 1992; Dweck & Elliott, 1983; Schunk, 1984), evaluation practices (MacIver, 1988), on *either* the antecedents or consequences of motivation. Studies that have attempted to synthesize or meta-analyze the results of many studies in which the relationship between some motivational variable(s) and learning or achievement were investigated, and more recent studies that have applied path analytic models to simultaneously measure the direct and indirect effects

of motivational variables on achievement, all come to a similar conclusion: The observed correlation between motivation and achievement ranges from .12 to about .33 (Fraser, Walberg, Welch, & Hattie, 1987; Garcia-Celay & Tapia, 1992; Helmke, 1989; Hembree, 1988; Uguroglu & Walberg, 1979), with a maximum of approximately 10% of variance in achievement being explained by motivational variables.

Two common educational practices that have been found to influence antecedents and achievement consequences of motivation are provision of external rewards or incentives (Cotton & Cook, 1982; Fowler & Clingman, 1977; Morgan, 1984; Schunk, 1983), and the type of achievement goals (goal orientations) that are set for students (Ames, 1992; Ames & Archer, 1988; Elliott & Dweck, 1988; Nicholls, 1984).

Goal Orientation and Achievement

Two contrasting goal orientations have received considerable attention in motivation research (Ames, 1992). The two types of goal orientation have been given different labels by different researchers: Dweck (1986) calls them learning-oriented and performance-oriented goals; Nicholls (1984) and Graham and Golan (1991) use the labels task-involved and ego-involved goals; Ames and Archer (1988) refer to them as mastery-focused and ability-focused goals. A learning-oriented or task-involved or mastery-oriented goal is one that encourages and emphasizes the goal of personal accomplishment or self-improvement, of engaging in and mastering a task for its own sake. A performance-oriented or ego-involved or ability-focused goal orientation, on the other hand, encourages and emphasizes the goal of proving one's ability relative to the ability of others, of maintaining positive judgments of one's ability, learning being a means to an end rather than an end in itself.

Each of these goal orientations can be induced by different learning task structures, such as emphasizing the development of understanding versus successful completion, or by varying evaluation conditions such as using criterion-referenced versus norm-referenced assessment (Ames, 1992). Specific motivational and achievement patterns have been linked to the salience of either ego-involved or task-involved goal orientations. According to Ames's (1992) extensive review of the literature on goal orientations and

motivation, research evidence suggests that a task-involved goal orientation is associated with "a wide range of motivation-related variables [including perceived self-efficacy, effort, persistence] that are conducive to positive achievement activity and that are necessary mediators of self-regulated learning" (p. 262). In contrast, ego-involved goal orientations have been associated with a pattern of motivation that includes avoidance of challenging tasks (Elliott & Dweck, 1988), use of superficial learning strategies such as memorization (Meece, Blumenfeld, & Hoyle, 1988), and a perception that success is a function of ability rather than effort (Dweck, 1986).

There is evidence that goal orientations interact with particular student characteristics to produce different performance outcomes. Nicholls (1984) reviews a number of studies that examined the interactive effects of goal orientation and perceived self-efficacy. Nicholls (1984) concludes that "compared to task involvement, ego involvement produces lower performance in low-perceived-ability individuals and equal or higher performance in high-perceived-ability individuals" (p. 341).

Most of the studies that have compared goal orientations have examined their effects on performance during classroom learning activities rather than at the time of test taking. One study by Brown and Walberg (1993) examined the effect of a goal orientation set at the time of test taking only. However, the goal orientation that Brown and Walberg set falls into neither the ego nor task-involved goal orientation categories. Instead, the goal orientation they established at the beginning of a test related to evaluating the students' teachers on the basis of the students' performance. The mean test score of students who were told that their test results would reflect on the performance of their teachers was .3 standard deviations above the mean score of students who received the standard instructions for the Iowa Tests of Basic Skills (1978).

Extrinsic Rewards and Achievement

Although external rewards have been linked to a decrease in subsequent interest in tasks similar to those for which rewards were offered (Weinberg, 1978), offering tangible rewards for successful performance on an academic task tends to result in short-term increased effort, perseverance and performance on the task (Bandura, 1977; Schunk, 1984). Often the effects of

rewards vary with circumstances such as quantity and type of reward, goal proximity, or initial level of interest (Cotton & Cook, 1982; Morgan, 1984). For example, Schunk (1984) found that linking a reward to a particular level of achievement resulted in higher performance than simply offering a reward for engagement in the task.

Intervening Cognitive Processing Variables

Regardless of the magnitude of the relationship between motivational variables and achievement, more and more researchers take the view that any effects of motivational antecedents on achievement are mediated by cognitive processing variables that reflect the amount and type of mental effort invested during the learning or assessment task (Salomon, 1983). Researchers such as Corno and Mandinach (1983), Pintrich and De Groot, (1990), Zimmerman and Martinez-Pons (1990), Graham and Golan (1991), and Boekarts (1988) have recently become focused on examining the relationships among (a) antecedents of motivation such as self-efficacy or attributions; (b) effort, as manifested in regulation and control of information processing; and (c) final achievement outcomes. Effortful performance appears to be driven by a set of higher-order/metacognitive/non-automatic processes that support the acquisition, retrieval and application of knowledge (Corno & Mandinach, 1983). While various labels have been given to the components of metacognition, it includes planning one's work, monitoring (checking) one's work, being aware of one's thought processes, and use of task-relevant strategies such as elaboration, or relating a new problem to something familiar, or distinguishing between important and irrelevant information. Learners who employ metacognitive strategies have been called "self-regulated learners" (Corno, 1986; Zimmerman, 1986).

The results of correlational studies indicate that use of metacognitive strategies (self-reported) is related to perceived self-efficacy (Zimmerman & Martinez-Pons, 1990), perceived mastery (task-involved) goal orientation (Ames & Archer, 1988), and classroom performance (Pintrich & De Groot, 1990). As yet, there appear to be no published studies that investigate the direct and indirect causal paths from motivational antecedents through use of metacognitive strategies to achievement.

Test Anxiety

Effort and the nature and extent of cognitive processing during test-taking are not necessarily a function of effort expended and cognitive processing during learning. For example, a student might invest great effort during a classroom learning activity, but invest little effort during a test because the consequences of performance on the test are not important; another student might invest minimum effort during the learning and instruction phase of education, but might become highly motivated at the point when his or her knowledge is being assessed, particularly if there are serious consequences attached to his or her performance on the test. The latter student may have difficulty since no amount of metacognitive strategy use can substitute for the lack of relevant subject-matter knowledge that may have resulted from a mindless approach during learning.

One variable that operates at test taking time is test anxiety. Its causes and effects have been the subject of considerable research. There are two components of anxiety, a worry component and an emotional component (Liebert & Morris, 1967; Morris, Davis, & Hutchings, 1991). Worry refers to the cognitive elements of the anxiety experience, such as negative expectations and cognitive concerns about oneself, the situation at hand, and its potential consequences. Emotionality refers to one's perception of the physiological-affective elements of the anxiety experience, that is, indications of autonomic arousal and unpleasant feeling states such as nervousness and tension.

Significant negative correlations between worry and test performance (but not between emotionality and test performance) appear to hold for actual examination scores (e.g., Sieber, O'Neil, & Tobias, 1977), course grades (Hembree, 1988), and Graduate Record Examinations (Powers, 1986), as well as laboratory studies. The majority of correlations between worry and test performance range from -.1 to -.4, with the average correlation being -.31 (Hembree, 1988).

One explanation of the negative effects of test anxiety on test performance is in terms of a reduction in cognitive processing capacity (Tobias, 1985; Wine, 1971). A large portion of the cognitive processing capacity of text-anxious people is engaged in worry, thereby limiting the cognitive space (working memory) available for metacognition and task-relevant information

processing. Therefore, students with high levels of worry might be engaging in less metacognitive activity. However, research to date has yielded inconsistent findings in relation to the hypothesis that students with high anxiety engage in less metacognitive activity (Pintrich & De Groot, 1990). The inconsistent findings are at least partly due to the fact that estimates of metacognitive activity are mostly based on students' own perceptions; hence, highly anxious students might perceive themselves to be expending more mental effort as they try to compensate for the reduction in cognitive capacity that has resulted from too much anxiety.

In general, there is a need for more studies to focus on the effects on test performance of motivational antecedents (not just anxiety) introduced at the time of test taking. Because the effects of any motivational antecedent or set of motivational antecedents on achievement are mediated by effort or cognitive engagement, which in turn are manifested in level and type of cognitive information processing, then any study that would try to investigate the effects of motivation on performance would have to measure these intervening variables.

Ethnic and Gender Differences in Motivation and Achievement

Ethnic and gender differences have been found in performance on NAEP mathematics tests (Mullis, Dossey, Owen, & Phillips, 1991). In general, Asian and White students outperform Latino and African-American students. Males outperform females on two mathematics content areas in grade 8 (measurement and estimation), but on all content areas in grade 12. This pattern of gender differences is consistent with wider research on gender differences in mathematics achievement (Benbow & Stanley, 1980).

Gender differences have also been found on motivational antecedents and consequences other than performance (Dweck, 1986). For example, Teideman and McMahon (1985) found that girls responded to more types of rewards than did boys. Zimmerman and Martinez-Pons (1990) found that girls reported using more metacognitive strategies, but had lower perceptions of their ability, than boys. Females also have higher levels of test anxiety than males (Wigfield & Eccles, 1989).

Ethnic differences in motivational variables have not received much attention to date. Hembree (1988) found that the test anxiety of Whites and African-Americans was similar in high school, but that Latinos were consistently more test anxious than Whites. There is a need for studies that would examine the differential effects of motivational conditions on the cognitive processing and performance of different ethnic groups and of males and females.

Patterns of Non-Response to Test Items

Recently two studies have focused on patterns of non-response to items in 1986 and 1990 NAEP tests (Koretz, Lewis, Burstein, & Skewes-Cox, 1992; Swinton, 1991). A distinction is made between number of items omitted (that is, skipped during a test) and number of items not reached (the point in the test at which a student stopped attempting items). In NAEP mathematics tests, the not-reached rates decreased from 1986 to 1990. Few gender differences in non-response were found and most of the apparent differences between White and minority students reflect proficiency differences. However, the results of these two studies suggest that the routine monitoring and reporting of non-response patterns is warranted, particularly in studies where the effects of motivational variables on test performance are investigated.

Conclusion

Because it is impossible to manipulate "motivation" directly, one is forced to manipulate some of its antecedents, that is, variables that appear to influence engagement in cognitive activity, which, in turn, influences performance. In the studies reported below, goal orientations and financial incentives were manipulated. The effects of various motivational conditions on students' performance on a subset of 1990 NAEP mathematics test items, on non-response patterns, and on the intervening variables of perceived metacognition, effort, ability, and worry were examined. Differential effects of the motivational conditions on test scores, on patterns of non-response and on the perceived effort, worry, ability, and metacognition of males and females, and of different ethnic groups, also were investigated.

STUDIES CONDUCTED BY USC/CRESST AND CSE/CRESST

Introduction

What follows is a detailed description of the pilot and main studies conducted to investigate the effects of different motivational conditions on the performance of 8th- and 12th-grade students on a subset of released items from the 1990 NAEP mathematics test. The differential effects of the motivational conditions on male and female students and on students of four different ethnic backgrounds were examined.

The studies are reported in the sequence in which they were conducted (although the financial incentives and goal orientation pilot studies were conducted simultaneously). For each study, the procedure is described, followed by a detailed presentation of results. An overview of the data analyses conducted and the organization of the results sections is presented below. ANOVA summary tables are in Appendix C. All tables except ANOVA summary tables are integrated within the text. Summaries of results are provided at the end of the section on a particular study. A final summary of all results is provided before the discussion section.

I. FOCUS GROUP "INCENTIVES" STUDY

As mentioned above, the role of motivation in students' standardized testing performance has recently received national attention (ETS, 1991; Shanker, 1990). Specifically, differences in student motivation to perform well on standardized tests have been cited as one reason why U.S. students perform worse than students from many other industrialized nations on international assessments such as the National Assessment for Educational Progress (NAEP). In response to this motivational explanation, this study examined the extent to which student motivation might be increased through offering students incentives in five areas: material rewards, recognition, comparisons, consequences, and feedback. We were also interested in whether incentives preference would differ by gender and ethnicity. The purpose of this study was to identify incentives to use in our experimental work. For each of the five incentive areas, subjects were presented with a base list of incentives and had five minutes to brainstorm additional ideas. For each area, the subjects were also instructed to write, on individual response

sheets, which of the listed incentives would motivate them the most and second most to do their best on a standardized test and which, if any, would discourage them. Finally, subjects were asked to select one incentive across all five incentive areas that would most motivate them to do their best and one that would most discourage them. Subjects listed material rewards such as college scholarships, class parties, and money as the most motivating incentives. However, they also listed some of the incentives from the other four categories as highly motivating. Moreover, the ranking of the incentives differed by grade level, SES, and ethnicity.

Method

Subjects

Eight groups of 8th-grade students and eight groups of 12th-grade students participated in this study. The group sizes ranged from six to eleven, making a total of 67 female and 68 male subjects. One male subject was omitted because of missing data. Each group made up one cell of a 2x2x4 design with grade level, socioeconomic status (SES), and ethnicity as the independent variables. Socioeconomic status consisted of two groups: low SES (determined by participation in school lunch programs) and high SES (determined by selecting schools in higher income neighborhoods.) Four ethnic/racial groups were represented: Whites, Asians, African-Americans, and Latinos (see Table 1).

Procedure

Participating schools were asked to assemble a gender-balanced group of eight 8th- or 12th-grade students of a particular ethnicity and socioeconomic status. Actual group size varied across school sites. Facilitators ran 1-hour focus groups in an available classroom or resource room at the school where the students were enrolled. School staff provided the subjects' grade point averages based on the subjects' transcripts. Schools were provided with a small honorarium for their participation.

Table 1

Distribution of Students in Focus Group Studies by Grade Level, SES, and Ethnicity

	12th grade		8th grade	
	High SES	Low SES	High SES	Low SES
White	n = 6	n = 11	n = 9	n = 10
Asian	n = 9	n = 7	n = 6	n = 8
African-American	n = 8	n = 11	n = 7	n = 9
Latino	n = 8	n = 9	n = 8	n = 8

Four female focus group facilitators, who were ethnically similar to the groups in this study, were trained using a transcript to lead their focus group members in 5 brainstorm sessions and to instruct their group members in filling out the individual response measures.

The facilitator of the focus groups began all focus group sessions by explaining the purpose of the study and giving all the subjects a chance to withdraw from the study. Only one subject chose not to participate. Next, the facilitator showed the subjects a *California Test of Basic Skills* (CTBS) booklet and made sure that all subjects clearly understood what school-wide standardized tests are and that they all had experience with taking them. Once the facilitator had conveyed to the subjects that all the remaining questions concerned standardized tests only, she asked the students to write down their answers to the three following questions: (a) How hard do you try on standardized tests? (on a scale where 1 equals *Not at all* and 4 equals *Really hard*); (b) Regardless of how hard you usually try, what would encourage you to do your best?; and (c) What discourages you from doing your best?

Within each focus group, a 5-minute brainstorm session occurred for each of the incentive areas being studied: material rewards, recognition, comparisons, consequences, and feedback. Subjects were instructed that when they brainstorm, they should come up with as many ideas as possible, not be critical of one's own or others' ideas, and try to be creative. Each of the

brainstorm sessions began with the presentation of a base list of incentives that students, such as themselves, would receive based on their performance on a standardized test. (The base lists were products of a research literature review and a brainstorm session held with some young college students and researchers with children the age of the study's subjects.) The focus group members were presented with each base list and given five minutes to brainstorm additional ideas for the list. As the group members made suggestions, their ideas were added to the original list.

At the end of the 5-minute brainstorm or when subjects no longer had any ideas to add to the list, the subjects were instructed to write down, on their individual response sheets, the incentives from the list just generated that would first and second most motivate them, as individuals, to work harder on a standardized test. In addition, they were told that if there was something on that list that would discourage them from trying to do their best—something that would make them try less hard—then they should write it down in the space provided. If there was not anything that would discourage them, they were told to leave that space blank. This ranking process occurred for the following seven categories: (a) material rewards for individual students; (b) material rewards for classes; (c) recognition for individual students; (d) recognition for classes; (e) comparisons made between students and groups of students; (f) academic and funding consequences; and (g) performance feedback. Finally, subjects were asked to write down any other ideas that had not been covered in any of the other lists that would motivate them to work harder on a standardized tests.

After the subjects completed their ranking by category, they were asked to select the one incentive across all the categories that would most motivate them to try their hardest on a standardized test and to circle that item on their response sheet. The subjects were also asked to underline the most discouraging item on their response sheet if they had listed more than one discouraging item.

Results

Open-ended Responses

Extent of student effort. The subjects were asked to indicate on a 4-point scale how hard they usually try to do their best on standardized tests. Even though most (61%) of the subjects said they try "pretty hard," only 22% of the subjects responded that they try "really hard"; 13% indicated that they tried a "little bit" and 3.5% said that they did not try at all. Compared to the ETS data (ETS, 1991), which is lower, this may indicate that students, when interviewed, state that they try harder than they do when asked by an anonymous survey.

What is encouraging. The subjects' open-ended responses to the question "What would encourage you to do your best on a standardized test?" primarily fell into five categories in the following order: (a) importance of the test; (b) self-satisfaction; (c) parent approval; (d) recognition for high performance; and (e) characteristics of the test. The most common responses that subjects gave were incentives that would make tests count or be important because they would affect the students' school records, school reputations, college admissions, grade point averages, grade advancements, academic tracks, permission to play sports, or futures in general (n = 35). Self-satisfaction, which includes doing one's best for one's self, was the second most popular response (n = 31). The next most common response that subjects made was to please their parents (n = 27), but this response overwhelmingly came from 8th-grade students as opposed to 12th-grade students. Fourth most commonly mentioned was some form of recognition for high performance, such as prizes, awards, praise, money, scholarships, or privileges (n = 16). Finally, characteristics of the test that might be improved, such as making the tests shorter or more interesting, were mentioned as incentives by a few subjects (n = 6). On open-ended responses, money was seldom mentioned.

What is discouraging. When asked "What most discourages you from doing your best on standardized tests?", the subjects' responses mainly fell into the following categories: (a) poor characteristics of the test; (b) unimportance of the test; (c) nothing; (d) pressure or nervousness caused by the test; and (e) physical and affective states at the time of taking the test. The most common response made by the subjects was that the long length, lack of variation of test items from year to year, and boring or confusing test content

discouraged them from doing their best on standardized tests (n = 45). The next most common response was the lack of importance of standardized tests, but this was mainly a concern of 12th-grade as opposed to 8th-grade subjects (n = 24). The unimportance of tests was exemplified by tests having no bearing on college admissions, not counting in general, and receiving little concern from teachers and other people. The third most common response of the subjects was that nothing discouraged them from doing their best (n = 14). The fourth most common response was nervousness about not doing well on the test (n = 12). Finally, a few subjects mentioned emotional or physical discomforts (e.g., being hungry, tired, hot, sick, or angry) that discourage them from trying their best on standardized tests (n = 3).

Ranking of Incentives

Overall most encouraging incentives. Except for a couple of incentives, there was little agreement over what one incentive would encourage students to try their hardest on a standardized test. A college scholarship was the overwhelming choice, followed by money. Paying for SAT fees or a college admission application, writing a letter of recommendation to a college of choice, and tickets for an amusement park were tied for third place. Except for the letter of recommendation and test scores affecting college admission, all of the high frequency incentives involved money (75/134 or 56%). Money in some form was seen as the most encouraging incentive by slightly more than one-half of the students. For all the remaining incentives named as most encouraging, there was little if any agreement (see Table 2).

Overall most discouraging incentives. When asked which one of the incentives discussed in the focus group might discourage then from trying their hardest on a standardized test, most of the subjects wrote no response or actually said "Nothing." Very little agreement exists between the subjects over what would be discouraging. Table 3 lists the most commonly mentioned incentives that students find to be discouraging and the frequency with which they were mentioned.

Within-Category Ranking and Demographic Differences

Within each domain of incentives studied, we compared the incentives' popularity by calculating a mean score for each incentive. Mean scores were

Table 2

Most Encouraging Incentives

Incentive	Frequency
College scholarship	35
Money	19
Recommendation to college of choice	10
Pay for SAT or college admission application	9
Tickets to an amusement park	6
Free movie tickets	3
Free prom tickets	3
Test scores affect college admission	3

Table 3

Most Discouraging Incentives

Incentive	Frequency
Nothing	45
Poor test performance can hold you back a grade	9
Individual student compared to other students	3
Comparisons between individual students	6
Free video arcade tokens	3
Teacher tells you that you did well	5
Individual students are compared by parents	4
Be able to get a face-to-face explanation on missed test items	4
Free yearbook	3

calculated by assigning a value of 3 to items ranked as first most motivating, a value of 2 to items ranked second most motivating, and a value of zero to items ranked as most discouraging or not mentioned among the subjects' rankings. Due to the wide array of incentives that subjects listed, we felt it necessary to limit our comparisons to those incentives with appeal to many subjects. Twenty-four high frequency incentives were selected for further investigation based on their mean score being greater than .40.

The following is a description of how the most commonly listed incentives ranked among their own category of incentives and how the popularity of those

incentives differed due to demographic differences in the following areas: grade level, ethnicity, gender, socioeconomic status, and grade point average. Demographic differences were determined by running ANOVAs on the incentives' means.

Material rewards. The subjects preferred class activities and money for college related fees or for whatever the student wanted when choosing what material rewards would motivate them to try their hardest on a standardized test. Class activities included: having a class party, going on a class or school field trip, or going with the class to a restaurant. When subjects mentioned money, the amount of money was often unspecified and when it was specified it ranged from 20 to 200 dollars. The college related fees included: college scholarships and paying for the students' SAT or college application fees (see Table 4).

How motivating some of these rewards were differed by grade level, ethnicity, SES, grade point average, and gender. The most popular reward, class party, was reported as significantly more encouraging to 8th-grade students (M = 1.59) than to 12th-grade students (M = .63) ($F(1, 133) = 21.27$, p = .001) as well as most encouraging to White (M = 1.53) subjects, followed by African-American (M = 1.25) and Latino (M = .94) subjects and lastly by Asian (M = .74) subjects. The latter effect approached statistical significance ($F(3, 131) = 2.41$, p = .07). The second most popular reward, money, was also reported to be more motivating by 8th graders (M = 1.44) as opposed to 12th

Table 4

Most Motivating Material Rewards, Their Means and Standard Deviations

	M	*SD*
Class Party	1.13	1.29
Money	1.01	1.31
Class or school field trip	.92	1.29
Scholarships for college	.89	1.42
Class Restaurant Trip	.61	1.10
Pay SAT or college admission application	.44	.96

graders (M = .53) ($F(1, 133) = 18.02$, $p < .01$) as well as by higher SES students (M = 1.32) as opposed to lower SES students (M = .63) ($F(1, 133) = 10.17$, $p = .001$). As might be expected, a college scholarship was a more motivating incentive for subjects with higher grade point averages (M = 1.23) than those with lower ones (M = .57) ($F(1, 133) = 7.59$, $p = .001$). A class restaurant trip was more motivating for African-American (M = 1.0) and White (M = .69) subjects than for Asian (M = .58) or Latino (M = .09) subjects ($F(3, 131) = 4.46$, $p = .001$). Finally, female subjects (M = .63) valued receiving fees for the SAT or a college application more than male subjects did (M = .25) ($F(1, 133) = 5.4$, $p = .02$).

Recognition. Personally appearing on television, as a form of recognition for doing well on a standardized test was the most popular incentive and appearing on television as a class was the third most popular incentive among 8th- and 12th-grade subjects alike. Second most popular was the suggestion that parents be sent a letter that recognizes students' high performances. The fourth and fifth most popular forms of recognition listed by the subjects were receiving a certificate or award as a class or as an individual. Finally, many subjects mentioned receiving a letter of recommendation to a college of their choice as a motivating incentive (see Table 5).

There were ethnic and SES differences regarding how motivating appearing on television would be. Asian subjects (M = .26) were less motivated by the prospect of appearing on television than African-American (M = 1.67),

Table 5

Most Motivating Forms of Recognition, Their Means and Standard Deviations

	M	*SD*
Student TV appearance	.98	1.35
Letter of recognition sent home to parents	.64	1.14
Class TV appearance	.69	1.21
Class certificate or award	.57	1.05
Receive Certificate or award	.56	1.09
Recommendations for colleges of choice	.53	1.11

White (M = 1.03), and Latino subjects (M = .79) ($F(1, 131) = 7.55$, p = .001), and higher SES subjects (M = .71) were also less motivated to be on television than lower SES subjects (M = 1.29) ($F(1, 133) = 6.4$, p = .01). The suggestion to send a letter of recognition home to students' parents was better received by 8th-grade students (M = .99) than 12th-grade students (M = .28) ($F(1, 133) = 16.93$, $p < .01$). Also, 8th-grade students (M = .86) preferred receiving a class certificate or award more than 12th-grade students (M = .26) did ($F(1, 133) = 11.95$, p = .001). Finally, grade level, ethnicity, SES, and grade point average were factors that influenced how much a subject was motivated by receiving a letter of recommendation to his or her college of choice. As might be expected, receiving a letter of recommendation was only mentioned by 12th-grade students (M = 1.10) ($F(1, 133) = 35.8$, p = .001). Although this suggestion was popular among Asian (M = 1.0), White (M = .48), and African-American subjects (M = .41), no Latino subjects (M = 0.) listed it as a motivating incentive ($F(3, 131) = 6.23$, p = .001). In addition, higher SES subjects (M = .79)and subjects with higher grade point averages (M = .85) were more motivated by receiving a letter of recommendation than lower SES subjects (M = .16) and subjects with lower grade point averages (M = .15), respectively ($F(1, 133) = 11.8, p < .01$) ($F(1, 133) = 14.4, p < .01$).

Comparisons. The three most motivating comparisons were school scores being compared to other school scores, individual students' scores being compared to other individual students' scores by teachers, and the average student score of the United States being compared to other countries' scores (see Table 6). Comparing schools' scores was more motivating to male subjects (M = .96) than to female subjects (M = .53) ($F(1, 133) = 4.24$, p = .04), and to Latino subjects (1.33) than to Asian (M = .61), African-American (M = .58), and White (M = .43) subjects ($F(3, 131) = 3.43$, p = .02). The comparison of individual student scores was more motivating to lower SES subjects (M = .32) than to higher SES subjects (M = .63) ($F(1, 133) = 3.17$, p = .08). Finally, comparing different countries' scores was much more motivating to 8th-grade subjects (M = .67) than to 12th-grade subjects (M = .20) ($F(1, 133) = 7.49, p < .01$), and to White subjects (M = 1.12) than to Latino (M = .49), Asian (M = .26), or African-American (M = 0.0) subjects ($F(3, 131) = 7.88, p < .01$).

Table 6

Most Motivating Comparisons, Their Means and Standard Deviations

	M	*SD*
School scores are compared to other school scores	.74	1.22
Individual students are compared to each other by the teacher	.49	1.03
Compare countries' scores	.44	1.00

Consequences. The suggestion that standardized test scores would count towards students' regular class grades was the most motivating consequence overall. The suggestion that poor test performance might keep you back a grade was the second most motivating consequence. The third most motivating consequence was the idea that schools would receive more funding if they performed better on standardized tests. Finally, the idea that parents would be sent test scores and rankings for their children regardless of how well the students performed was also seen as a motivating consequence by many of the subjects (see Table 7).

For two of the popularly mentioned consequences, demographic differences existed. African-American subjects (M = 1.43) reported being more motivated by the suggestion that poor test performance can keep a student back a grade than White (M = .88), Latino (M = .70), or Asian (M = .48) subjects ($F(3, 131) = .98$, $p = .01$). The consequence that parents will be sent test scores was reported as more motivating by 8th-grade subjects (M = 1.0) than by 12th-grade

Table 7

Most Motivating Consequences, Their Means and Standard Deviations

	M	*SD*
Test performance counts toward regular class grade	.90	1.26
Poor test performance can keep you back a grade	.90	1.29
Better class test performance gets more school funding	.69	1.14
Parents are sent scores and ranking	.64	1.11

subjects (M = .20) ($F(1, 133 = 23.9, p < .01$) and as more motivating to subjects with lower grade point averages (M = .84) than to those with higher ones (M = .42) ($F(1, 133) = 5.8, p = .02$).

Feedback. Five types of feedback regarding test performance were commonly mentioned as something that would motivate the subjects to try harder on standardized tests. They were: (a) receiving back information on one's strengths and weaknesses; (b) receiving explanations and correct answers for missed test items; (c) being able to get a face-to-face explanation on missed test items; (d) receiving back a test score or ranking in the class; and (e) receiving back correct items for missed test items (see Table 8).

Once again, there were differences in how motivating these forms of feedback were perceived related to subject, gender, and socioeconomic status. First, female subjects (M = 1.14) ranked receiving explanations and correct answers for missed test items higher than did male subjects (M = .59) ($F(1, 133) = 6.5, p = .01$) while male subjects (M = .88) ranked receiving only the test score or their ranking in the class higher than female subjects (M = .46) ($F(1, 133) = 5.00, p = .03$). Finally, lower SES subjects (M = .77) reported that they would be more motivated by receiving correct answers than the higher SES subjects (M = .33) reported ($F(1, 133) = 6.65, p = .01$).

Implications for Incentives Used in Pilot Studies

Since financial incentives were ranked high by both 8th- and 12th-grade students as motivators to try hard on tests, it was decided to conduct a pilot study that would compare the relative effectiveness of different financial incentives.

Table 8

Most Motivating Forms of Feedback, Their Means and Standard Deviations

	M	*SD*
Receive back information on your strengths and weaknesses	1.06	1.30
Get explanations and correct answers for missed test items	.87	1.27
Be able to get a face-to-face explanation on missed test items	.88	1.25
Receive back score alone or ranking in the class	.67	1.12
Receive back correct items for missed test items	.53	1.07

II. PILOT STUDIES

A. Financial Incentives Pilot Studies

Two financial incentives pilot studies were conducted. The first study compared the mathematics performance of four groups of 8th- and 12th-grade students; mathematics performance was measured using two blocks (3 and 7) of released items from the 1990 NAEP mathematics assessment for grades 8 and 12. Each subject received either one of three different financial incentives—50 cents for every item answered correctly; $1 for every item beyond 8 items answered correctly (approximately chance response rate); a reward of $16 if the average score in the class was at least 24—or the standard NAEP instructions for two blocks of the NAEP mathematics test. Half of the students in each treatment group received the easier block of mathematics items (Block 3) prior to the more difficult block (Block 7), and half received the difficult set prior to the easier set.

The second financial incentives study was like the first except that only 12th-grade subjects were used and they were given the mathematics test items appropriate for 8th grade. It was reasoned that the motivational effects of the incentives might be more apparent on an "easier" test. The more relevant knowledge a student has, the more likely it is that increased effort will result in increased performance.

Procedure (Financial Incentives Pilot Studies)

Subjects and Assignment to Treatment Groups

Study 1 (8th and 12th grade). One hundred and sixty-six 8th-grade students and 215 12th-grade students from 4 schools in Southern California were tested. Schools were selected to provide a range of socioeconomic and ethnic backgrounds. An honorarium of $75 per class was paid to each school that participated. Table 9 shows the ethnic breakdown of the subjects.

The numbers of males and females in the sample are summarized in Table 10.

Table 9

Financial Incentives Pilot Study 1: Ethnic Breakdown of Sample by Grade

Ethnic group	8th grade	12th grade
White	42	70
African-American	92	63
Latino	24	67
Asian	5	15
Other	3	0
Total	166	215

Table 10

Financial Incentives Pilot Study 1: Gender Breakdown of Sample by Grade

Gender	8th grade	12th grade
Male	84	110
Female	82	105
Total	166	215

For each grade level, subjects within each of five ethnic groupings (White, African-American, Latino, Asian, and Other) were randomly assigned (across schools) to 8 treatment conditions. There were 8 treatment conditions because the order of the easy and difficult mathematics blocks was varied within each treatment. The numbers of students assigned to each condition are displayed in Table 11 (numbers in cells within each grade level are not equal because some subjects who were initially assigned to treatment conditions were absent from school on the day that the test was administered).

Table 11

Financial Incentives Pilot Study 1: Number of Subjects by Treatment by Grade

Treatment condition	8th grade	12th grade
50 cents, Easy first	20	26
50 cents, Difficult first	25	28
$1 after 8, Easy first	18	25
$1 after 8, Difficult first	19	28
$16, class mean 24, Easy first	19	27
$16, class mean 24, Difficult first	18	27
Control, Easy first	25	27
Control, Difficult first	22	27
Total	166	215

Since a two-way ANOVA with treatment and order as independent variables indicated neither a main nor an interaction effect of order, for subsequent analysis purposes, the number of treatments was reduced to four, reflecting the three experimental motivation conditions and the control condition. Because there were so few Asians and students in the "Other" ethnic category, they were not included in the analysis. This left a total of 158 8th-grade and 200 12th-grade students for whom data were analyzed. Tables 12 and 13 below show the final number of subjects in each cell of the treatment by ethnicity by gender design.

Study 2 (12th-grade subjects given 8th-grade mathematics test). Two hundred and eleven 12th-grade students in two schools in Southern California received the 8th-grade mathematics test. The ethnic and gender breakdown of that sample and the numbers in each treatment condition are shown in Tables 14, 15 and 16. Within each ethnic group, subjects were randomly assigned (across schools) to 8 treatment conditions.

Table 12

Financial Incentives Pilot Study 1, 8th Grade
Number of Subjects Tested by Treatment, Ethnicity and Gender (N=158)

Treatment group	Ethnicity											
	White			African-American			Latino			Total		
	M	F	All	M	F	All	M	F	All	M	F	All
50 cents	7	3	10	11	13	24	2	3	5	20	19	39
$1.00 after 8	4	6	10	10	12	22	3	2	5	17	20	37
Class	4	6	10	12	9	21	3	2	5	19	17	36
Control	7	5	12	10	15	25	6	3	9	23	23	46
Total	22	20	42	43	49	92	14	10	24	79	79	158

Table 13

Financial Incentives Pilot Study 1, 12th Grade
Number of Subjects Tested by Treatment, Ethnicity and Gender (N=200)

Treatment group	Ethnicity											
	White			African-American			Latino			Total		
	M	F	All	M	F	All	M	F	All	M	F	All
50 cents	8	8	16	6	10	16	9	9	18	23	27	50
$1.00 after 8	10	8	18	8	8	16	10	5	15	28	21	49
Class	8	12	20	10	5	15	8	7	15	26	24	50
Control	6	10	16	7	9	16	10	9	19	23	28	51
Total	32	38	70	31	32	63	37	30	67	100	100	200

Table 14

Financial Incentives Pilot Study 2: Ethnic Breakdown of Sample

Ethnic group	12th grade
White	108
African-American	23
Latino	62
Asian	16
Other	2
Total	211

Table 15

Financial Incentives Pilot Study 2: Gender Breakdown of Sample

Gender	12th grade
Male	112
Female	99
Total	211

Table 16

Financial Incentives Pilot Study 2: Treatment Breakdown of Sample

Treatment condition	12th grade
50 cents, Easy first	23
50 cents, Difficult first	33
$1 after 8, Easy first	29
$1 after 8, Difficult first	23
$16, class mean 24, Easy first	23
$16, class mean 24, Difficult first	24
Control, Easy first	28
Control, Difficult first	28
Total	211

Since the results on an ANOVA indicated that neither the main effect nor interaction of order with treatment were significant, for subsequent analysis purposes, the number of treatments was reduced to four, reflecting the three experimental motivation conditions and the control group. Because there were so few African-Americans, Asians and students in the "Other" ethnic category, they were not included in the analysis. This left a total of 170 12th-grade students for whom data were analyzed. Table 17 shows the final number of subjects in each cell of the treatment by ethnicity by gender design.

Materials and administration. In both financial incentives studies, each student received a booklet which contained two blocks of mathematics items from the 1990 NAEP mathematics test and a self-assessment questionnaire that consisted of 53 items. Fifty-one of the items represented four metacognitive variables (perceived planning, self-checking, cognitive strategy use, and awareness), as well as perceived effort, curiosity, and worry. The final two questions asked students to report their average grade in mathematics at the end of the previous semester and to rank their mathematics ability compared to their classmates. The history of the development of the self-assessment questionnaire is described in Appendix A.

Table 17

Financial Incentives Pilot Study 2, Grade 12: Number of Subjects Tested by Treatment, Ethnicity and Gender (N=170)

Treatment group	Ethnicity								
	White			Latino			Total		
	M	F	All	M	F	All	M	F	All
50 cents	13	15	28	5	10	15	18	25	43
$1.00 after 8	13	17	30	5	8	13	18	25	43
Class	15	10	25	9	5	14	24	15	39
Control	19	6	25	13	7	20	32	13	45
Total	60	48	108	32	30	62	92	78	170

A standard script was developed for administration of the test booklets (see Appendix B). A group of 14 retired school personnel and one graduate student were recruited and trained to administer the test booklets. These 15 test administrators were used in all of the studies. The ethnic breakdown of the test administrators was: 7 White, 6 African-American, and 2 Asian. The booklets were administered during one regular class period. The length of class periods ranged from 45 to 55 minutes. Students tested in the shorter class periods were less likely to complete all items in the self-assessment questionnaire, since that was the last part of the booklet. In each school, administrations were sequenced during the school day; therefore some classes were tested before others.

Scoring of open-ended items. In the 8th-grade mathematics test there were five open-ended items, and in the 12th-grade test there were eight. These were scored by three raters according to the NAEP 1990 scoring system. The raters were graduate students who had taught mathematics at the secondary school level. For all pilot studies, interrater agreement for the 8th-grade items ranged from 91% to 100%, and for the 12th-grade items, from 95% to 100%.

Follow-up with students. Approximately one month following data collection, the persons who had administered the tests went back to each school with a letter for each student. The letter contained information about the student's score on the mathematics items, the 25th and 50th percentile scores on those items based on the 1990 NAEP data, and the appropriate amount of money in the form of cash. All students in the control groups received a $5 payment (which they were not expecting).

Analyses Conducted on Data From Pilot Studies (Financial Incentives and Goal Orientation Pilot Studies)

For each experimental pilot study, four analyses (described below) were conducted. For Analyses 1, 2 and 3, only ethnic groups with reasonable numbers of subjects were included. The final numbers of subjects for analysis are presented in Table 18.

Table 18

Number of Subjects Used for Analysis in Pilot Studies

Study	Grade	Ethnic groups included	Total N
Financial incentives, Pilot 1	8	White, African-American, Latino	158
	12	White, African-American, Latino	200
Financial incentives, Pilot 2	12	White, Latino	170
Goal-orientation Pilot	8	White, African-American	173
	12	White, Latino	197

Test administrators noticed that where the test had already been administered in a school, some subjects in classes subsequently tested were aware of the nature of the study and differences between test instructions. Because of concerns for contamination of treatment effects, it was decided to perform additional analyses using only the data from students tested first in all schools; these additional analyses are described in Analysis 4 below.

Analysis 1. Univariate analysis of variance. Seven mathematics achievement variables, three non-response variables, four affective variables, and one other variable (self-reported previous mathematics achievement) were treated as dependent variables in completely randomized factorial ANOVAs with treatment group, ethnicity, and gender as the independent factors. The seven mathematics achievement variables were:

1. total score on the test (Block 3 and Block 7);
2. score on Block 3 test items;
3. score on Block 7 test items;
4. score on "easy" items, defined as items that at least 75% of students in the 1990 NAEP National Sample answered correctly (8th grade: 9 items; 12th grade: 10 items);

5. score on "moderately difficult" items, defined as items that between 48% and 65% of students in the 1990 NAEP National Sample answered correctly (8th grade and 12th grade: 10 items);

6. score on "difficult" items, defined as items that less than 40% of students in the 1990 NAEP National Sample answered correctly (8th grade: 12 items; 12th grade: 9 items);

7. score on open-ended items (8th grade: 5 items; 12th grade: 8 items).

The three non-response variables were:

1. sum of the number of items skipped in each block of items;

2. sum of the number of items not reached at the end of Block 3 and Block 7;

3. number of items not attempted in the test, defined as the sum of the number of items skipped and the number of items not reached in each block.

The four affective variables were:

1. perceived worry, defined as score on the worry scale that was part of the self-assessment questionnaire;

2. perceived effort, defined as score on the effort scale that was part of the self-assessment questionnaire;

3. perceived curiosity, defined as score on the curiosity scale that was part of the self-assessment questionnaire;

4. perceived mathematics ability, defined as students' ranking of their mathematics ability compared to their classmates (*much less than most*, *less than most*, *equal to most*, *better than most*, or *much better than most*).

Separate ANOVAs rather than MANOVA analyses were conducted because six of the seven mathematics achievement variables were subsets of the total mathematics score, two of the three non-response variables were subsets of the third, and the affective variables are theoretically separate.

Since, for most of the *F*-tests of significance of differences among groups, cell frequencies were unequal, the unique effect of each independent variable and interaction was tested using the "regression" approach for decomposing sums of squares (Winer, Brown, & Michels, 1991). In some cases, the variances of the groups being compared were unequal. However, the *F*-test is

robust to violations of assumptions, even in unbalanced designs (Abedi, 1974). Simple main effect and Scheffe post hoc comparison analyses were conducted when significant interaction or main effects were found.

Analysis 2. Multivariate analysis of variance. Four metacognitive variables (perceived planning, perceived awareness, perceived self-checking, and perceived cognitive strategy use) were combined in one MANOVA because those four variables reflect a common construct called metacognition. Treatment, ethnicity, and gender were the independent variables. Whenever a multivariate F-test revealed a significant effect of some independent variable(s) on the combined metacognitive variables, then post hoc univariate F-tests were examined to ascertain which of the dependent variables contributed most to the differences among the groups. Significant univariate F-ratios were followed up with tests of simple main effects and/or Scheffe multiple comparison tests as appropriate. If univariate F-ratios were not significant (but the multivariate F-ratio was significant), then raw discriminant function coefficients were used to create a new "metacognition" variable which was a linear combination of the four separate variables. The significance of differences among the means on this new variable were then tested using the Scheffe post hoc comparison procedure.

Analysis 3. Correlations. The correlations between total mathematics score and each of the metacognitive and affective variables were examined.

Analysis 4. Subsample of subjects tested first in all schools. Data from those students tested first in schools were analyzed with treatment as the only independent variable, and each of the mathematics achievement, non-response, affective, and metacognitive variables being treated as dependent variables in ANOVA and MANOVA analyses.

Presentation of Results (All Pilot Studies)

In the following presentation of results, descriptions of analysis of variance results are limited to those where significant F-ratios were found. Unless there was a significant effect on a mathematics achievement variable (other than total score) that was different from the effects on total mathematics score, only effects on total mathematics score are discussed. The results of the simple main effects and Scheffe post hoc comparison analyses are reported in the text where appropriate (that is, whenever an overall F-test was

significant). Reference is made to detailed ANOVA tables, which are included in Appendix C of this report, and to tables of means and standard deviations, which appear throughout the text. Detailed descriptions of results for 8th- and 12th-grade samples are followed by a summary of results. A discussion of the results of the pilot studies precedes the report of the main study. For each study, the order of presentation of results is as follows:

1. ANOVA and MANOVA Results, 8th Grade
    - A. Treatment Effects
        - A.1 Treatment effects on mathematics achievement variables (including any interactions between treatment and ethnicity or gender)
        - A.2 Treatment effects on non-response variables (including interactions with ethnicity and gender)
        - A.3 Treatment effects on metacognitive and affective variables (including interactions with ethnicity and gender)
    - B. Ethnic Differences
        - B.1 Ethnic differences in mathematics achievement variables (main effect only)
        - B.2 Ethnic differences in non-response variables (main effect only)
        - B.3 Ethnic differences in metacognitive and affective variables (main effect only)
    - C. Gender Differences
        - C.1 Gender differences in mathematics achievement variables (main effect only)
        - C.2 Gender differences in non-response variables (main effect only)
        - C.3 Gender differences in metacognitive and affective variables
2. ANOVA and MANOVA Results, 12th Grade
    - A. Treatment Effects
        - A.1 Treatment effects on mathematics achievement variables (including any interactions between treatment and ethnicity or gender)

A.2 Treatment effects on non-response variables (including interactions with ethnicity and gender)

A.3 Treatment effects on metacognitive and affective variables (including interactions with ethnicity and gender)

B. Ethnic Differences

B.1 Ethnic differences in mathematics achievement variables (main effect only)

B.2 Ethnic differences in non-response variables (main effect only)

B.3 Ethnic differences in metacognitive and affective variables (main effect only)

C. Gender Differences

C.1 Gender differences in mathematics achievement variables (main effect only)

C.2 Gender differences in non-response variables (main effect only)

C.3 Gender differences in metacognitive and affective variables

3. Correlations, 8th Grade

Correlations between total mathematics score and metacognitive and affective variables

4. Correlations, 12th Grade

Correlations between total mathematics score and metacognitive and affective variables

5. Summary of Results

Results: Financial Incentives, Pilot Study 1 (8th grade, N=158; and 12th grade, N=200, Whites, African-Americans, and Latinos)

1. Univariate and Multivariate Analysis of Variance Results, 8th Grade

1.A. Treatment Effects

1.A.1. Treatment effects on mathematics achievement. There was no treatment effect of financial incentives on total mathematics score (see Table A1 in Appendix C), but treatment affected scores on moderately difficult items, $F(3, 134) = 3.8$, $p = .012$ (see Table A2 in Appendix C). However, post hoc Scheffe multiple comparisons did not reveal any significant differences between the scores of the treatment groups on moderately difficult items (see Table 19).

1.A.2. Treatment effects on non-response. Treatment interacted with ethnicity in its effect on number of items omitted, $F(6, 134) = 8.7$, $p < .001$, and number of items not attempted, $F(6, 134) = 3.0$, $p = .009$ (see Tables A3 and A4 in Appendix C). Analysis of simple main effects indicated that treatment affected the non-response of Latino students only. Scheffe post hoc multiple comparisons indicated that Latino students who were offered a financial reward based on the performance of their entire class attempted more test items than Latinos who received any other test instructions (see Tables 20 and 21). However, because the number of Latinos in this study was very small, this result should be interpreted with caution.

Table 19

Financial Incentives Pilot Study 1, Grade 8: Descriptive Statistics for Moderately Difficult Mathematics Items by Treatment (N=158)

Treatment	n	$\bar{X}$	SD
50 cents	39	6.3	3.1
$1.00 after 8	37	5.8	2.8
Class	36	5.0	2.8
Control	46	6.2	2.4
Total	158	5.8	2.8

Table 20

Descriptive Statistics for Number of Mathematics Items Omitted by Treatment and Ethnicity (N=158)

	Ethnicity											
	White			African-American			Latino			Total		
Treatment	n	$\bar{X}$	SD	n	$\bar{X}$	SD	n	$\bar{X}$	SD	n	$\bar{X}$	SD
50 cents	10	.2	.4	24	.7	.9	5	.4	.5	39	.5	.8
$1.00 after 8	10	.7	1.6	22	.5	.9	5	.2	.4	37	.5	1.1
Class	10	.7	1.2	21	.6	.8	5	4.8	4.5	36	1.2	2.3
Control	12	.2	.4	25	.6	1.7	9	.4	.5	46	.5	1.3
Total	42	.4	1.0	92	.6	1.1	24	1.3	2.7	158	.7	1.5

1.A.3. Treatment effects on metacognitive and affective variables. MANOVA revealed that treatment interacted with ethnicity in its effect on the combined metacognitive variables, multivariate $F(24, 378) = 2.05$, $p = .003$. Follow-up invariate F tests revealed a significant interaction effect on perceived self-checking (see Table A5 in Appendix C). Analysis of simple main effects indicated that treatment affected the perceived self-checking of Latino students only. Scheffe post hoc comparisons indicated that Latino students who were offered a financial reward based on the performance of their entire class reported less self-checking than Latinos who received any other test instructions (see Table 22). However, because the number of Latinos in this study was very small, this result should be interpreted with caution. There were no differences among groups on affective variables.

Table 21

Descriptive Statistics for Number of Mathematics Items Not Attempted by Treatment and Ethnicity (N=158)

	White			African-American			Latino			All		
Treatment	n	$\bar{X}$	SD	n	$\bar{X}$	SD	n	$\bar{X}$	SD	n	$\bar{X}$	SD
50 cents	10	1.7	2.6	24	2.5	3.8	5	.4	.5	39	2.0	3.3
$1.00 after 8	10	1.4	2.8	22	2.0	2.6	5	1.2	2.7	37	1.8	2.6
Class	10	2.7	2.6	21	2.1	2.4	5	7.6	7.1	36	3.1	3.8
Control	12	1.1	1.5	25	1.4	2.3	9	.8	1.0	46	1.2	1.9
Total	42	1.7	2.4	92	2.0	2.8	24	2.2	4.3	158	2.0	3.0

Table 22

Descriptive Statistics for Perceived Self-checking by Treatment and Ethnicity (N=158)

	White			African-American			Latino			All		
Treatment	n	$\bar{X}$	SD	n	$\bar{X}$	SD	n	$\bar{X}$	SD	n	$\bar{X}$	SD
50 cents	10	2.5	.51	23	2.4	.59	5	2.9	.5	38	2.5	.6
$1.00 after 8	10	2.9	.84	22	2.3	.68	5	2.9	2.7	37	2.5	.7
Class	10	2.6	.54	21	2.4	.65	5	1.7	7.1	36	2.4	.7
Control	12	2.4	.61	25	2.4	.74	9	2.4	1.0	46	2.4	.7
Total	42	2.6	.64	91	2.4	.66	24	2.5	4.3	157	2.4	.7

1.B. Ethnic Differences

1.B.1. Ethnic differences in mathematics achievement. ANOVA revealed a significant difference among ethnic groups on total mathematics scores, $F(2, 134) = 17.7$, $p < .001$ (see Table A1 in Appendix C). Scheffe post hoc comparisons revealed that Whites had higher total mathematics test scores (mean score = 24.6) than either African-Americans (mean = 17.6) or Latinos (mean score = 19) as evident in Table 23 below.

1.B.2. Ethnic differences in non-response. ANOVA revealed a significant difference among ethnic groups on number of items omitted, $F(2, 134) = 3.9$, $p = .022$ (see Table A3 in Appendix C). Latinos omitted more items than either Whites or African-Americans (see Table 24 below).

Table 23

Financial Incentives Pilot Study 1, Grade 8: Descriptive Statistics for Total Mathematics Score by Ethnicity (N=158)

Ethnicity	n	$\bar{X}$	SD
White	42	24.6	6.4
African-American	92	17.6	6.4
Latino	24	19.0	5.3
Total	158	19.7	6.9

Table 24

Financial Incentives Pilot Study 1, Grade 8: Descriptive Statistics for Number of Mathematics Items Omitted by Ethnicity (N=158)

Ethnicity	n	$\bar{X}$	SD
White	42	.4	1.0
African-American	92	.6	1.1
Latino	24	1.3	2.7
Total	158	.7	1.5

1.B.3. Ethnic differences in metacognitive and affective variables. Although intercorrelations among the four metacognitive variables ranged from .57 to .68, theory and previous research (Corno, 1986) have established these variables as separate constructs which may be differentially affected by treatments; therefore, the metacognitive variables were treated as four dependent variables in a multivariate analysis of variance. MANOVA revealed a significant ethnic difference on the combined metacognitive variables, multivariate $F(8, 216) = 2.67$, $p = .008$. Post hoc ANOVAs were not significant for any of the four metacognitive variables. Analysis of scores on a variable representing a linear combination of the four metacognitive variables indicated that Whites (mean = 1.3) and Latinos (mean = 1.1) reported more metacognitive activity than African-Americans (mean = .4). The raw discriminant function coefficients used to form the linear combination were .99 (perceived cognitive strategy use), 1.33 (perceived self-checking), -2.99 (perceived planning), and .57 (perceived awareness). African-Americans reported investing less effort than Whites (see Table 25 below and Table A6 in Appendix C).

1.C. Gender Differences

1.C.1. Gender differences in mathematics achievement. Males had higher scores (mean score = 14.3) than females (mean score = 12.9) on Block 3 mathematics items, $F(1,134) = 4.6$, $p = .034$ (see Table A7 in Appendix C, and Table 26 below).

Table 25

Financial Incentives Pilot Study 1, Grade 8: Descriptive Statistics for Effort by Ethnicity (N=156)

Ethnicity	n	$\bar{X}$	SD
White	42	3.4	.65
African-American	91	3.1	.66
Latino	23	3.3	.71
Total	156	3.2	.67

Table 26

Financial Incentives Pilot Study 1, Grade 8: Descriptive Statistics for Mathematics Block 3 by Gender (N=158)

Gender	n	$\bar{X}$	SD
Male	79	14.3	4.3
Female	79	12.9	4.6
Total	158	13.6	4.5

1.C.2. Gender differences in non-response. There were no differences between males and females in number of items omitted, number of items not reached, and number of items not attempted.

1.C.3. Gender differences in metacognitive and affective variables. Although intercorrelations among the four metacognitive variables ranged from .57 to .68, theory and previous research have established these variables as separate constructs which may be differentially affected by treatments; therefore, the metacognitive variables were treated as four dependent variables in a multivariate analysis of variance. Males and females did not differ on combined metacognitive variables or on the affective variables.

2. Univariate and Multivariate Analysis of Variance Results, 12th Grade

2.A. Treatment Effects

2.A.1. Treatment effects on mathematics achievement. There were no differences among the mathematics test scores of the four treatment groups.

2.A.2. Treatment effects on non-response. There were no differences among the treatment groups in terms of non-response to test items.

2.A.3. Treatment effects on metacognitive and affective variables. MANOVA revealed a significant difference among treatment groups on the combined metacognitive variables, multivariate $F(12, 437) = .88$, $p = .044$. However, follow-up ANOVAs were not significant. Analysis of scores on a variable representing a linear combination of the four metacognitive variables

indicated that students who were offered 50 cents per each item they answered correctly (mean = -.9) reported more metacognitive activity than did students who were offered $16 based on the average score of their class. The raw discriminant function coefficients used to form the linear combination of metacognitive variables were -2.18 (perceived cognitive strategy use), -1.11 (perceived self-checking), 2.23 (perceived planning), and .67 (perceived awareness).

2.B. Ethnic Differences

2.B.1. Ethnic differences in mathematics achievement. ANOVA revealed a significant difference among ethnic groups on total mathematics score, $F(2, 176) = 23.1$, $p < .001$ (see Table A8 in Appendix C). Scheffe post hoc comparisons revealed that Whites had higher total mathematics test scores (mean score = 29.1) than either African-Americans (mean = 20.9) or Latinos (mean score = 20.8), as presented in Table 27 below.

2.B.2. Ethnic differences in non-response. ANOVA revealed a significant difference among ethnic groups on number of items not reached, $F(2, 176) = 5.9$, $p = .003$, and on number of items not attempted, $F(2, 176) = 6.4$, $p = .002$ (see Tables A9 and A10 in Appendix C). African-Americans reached fewer items (that is, got less far in each block of test items) and attempted fewer items than did Whites (see Tables 28 and 29 below).

Table 27

Financial Incentives Pilot Study 1, Grade 12: Descriptive Statistics for Total Mathematics Score by Ethnicity (N=200)

Ethnicity	n	$\bar{X}$	SD
White	70	29.1	7.8
African-American	63	20.9	7.1
Latino	67	20.8	9.0
Total	200	23.7	8.9

Table 28

Financial Incentives Pilot Study 1, Grade 12: Descriptive Statistics for Number of Mathematics Items Not Reached by Ethnicity (N=200)

Ethnicity	n	$\bar{X}$	SD
White	70	1.9	2.7
African-American	63	3.9	3.6
Latino	67	3.6	3.8
Total	200	3.0	3.5

Table 29

Financial Incentives Pilot Study 1, Grade 12: Descriptive Statistics for Number of Mathematics Items Not Attempted by Ethnicity (N=200)

Ethnicity	n	$\bar{X}$	SD
White	70	3.0	3.2
African-American	63	5.4	3.9
Latino	67	4.8	4.5
Total	200	4.3	4.0

2.B.3. Ethnic differences in metacognitive and affective variables. There were no differences among ethnic groups on the combined metacognitive variables. Perceived mathematics ability, $F(2, 112) = 4.3$, $p = .016$, and worry, $F(2, 172) = 7.1$, $p = .001$, varied with ethnicity (see Tables A11 and A12 in Appendix C). Scheffe post hoc comparisons revealed that Latinos reported worrying more than White students (see Table 30 below). None of the group differences on perceived mathematics ability were significant in the Scheffe post hoc comparisons.

Table 30

Financial Incentives Pilot Study 1, Grade 12: Descriptive Statistics for Worry by Ethnicity (N=196)

Ethnicity	n	$\overline{X}$	SD
White	69	1.5	.49
African-American	63	1.8	.79
Latino	64	1.9	.82
Total	196	1.7	.73

2.C. Gender Differences

2.C.1. Gender differences in mathematics achievement. There were no differences between the mathematics test scores of male and female students.

2.C.2. Gender differences in non-response. Males omitted less items than females, $F(1, 176) = 5.8$, $p = .017$ (see Table A12a in Appendix C and Table 31 below).

2.C.3. Gender differences in metacognitive and affective variables. The only gender difference in metacognitive and affective variables was in worry. Females reported worrying more than males, $F(1, 172) = 6.9$, $p = .01$ (see Table A12b in Appendix C and Table 32 below).

Table 31

Financial Incentives Pilot Study 1, Grade 12: Descriptive Statistics for Number of Mathematics Items Omitted by Gender (N=200)

Gender	n	$\overline{X}$	SD
Male	100	1.0	1.5
Female	100	1.6	1.9
Total	200	1.3	1.7

Table 32

Financial Incentives Pilot Study 1, Grade 12: Descriptive Statistics for Worry by Gender (N=196)

Gender	n	$\bar{X}$	SD
Male	97	1.6	.66
Female	99	1.8	.79
Total	196	1.7	.73

3. Correlations, 8th Grade

Table 33 below shows the correlations between total mathematics test score and each metacognitive and affective variable. Total mathematics score was significantly positively correlated with cognitive strategy use (r = .18), previous mathematics grades (r = .32), and worry (r = -.27).

4. Correlations, 12th Grade

Total mathematics score was significantly positively correlated with worry (r = -.52), previous mathematics grades (r = .26), and perceived mathematics ability (r = .51) (see Table 34 below).

Table 33

Financial Incentives Pilot Study 1, Grade 8: Correlations Between Total Mathematics Score and Metacognitive/Affective Variables (Ns indicated in parentheses)

	A	CS	P	SC	W	E	C	PMG	PMA
Math Total	.09 (135)	.18* (157)	-.06 (156)	.05 (157)	-.27** (146)	.13 (156)	-.01 (153)	.32** (74)	.16 (71)

Note. A = Awareness; CS = Cognitive strategy use; P = Planning; SC = Self-checking; W = Worry; E = Effort; C = Curiosity; PMG = Previous mathematics grades; PMA = Perceived mathematics ability.

*p<.05. **p<.01 (two-tailed).

Table 34

Financial Incentives Pilot Study 1, Grade 12: Correlations Between Total Mathematics Score and Metacognitive/Affective Variables (Ns indicated in parentheses)

	A	CS	P	SC	W	E	C	PMG	PMA
Math Total	.11 (192)	.13 (198)	.04 (198)	.04 (198)	-.52* (196)	.09 (198)	-.02 (197)	.26* (138)	.51* (136)

Note. A = Awareness; CS = Cognitive strategy use; P = Planning; SC = Self-checking; W = Worry; E = Effort; C = Curiosity; PMG = Previous mathematics grades; PMA = Perceived mathematics ability.

*$p<.01$ (two-tailed).

5. Summary of Results: Financial Incentives Pilot Study 1 (8th grade, N=158; and 12th grade, N=200)

1. The only noteworthy effect of treatment occurred in grade 12. Students who were offered 50 cents per correct item reported engaging in more metacognitive activity than did students who were offered $16 based on the average score of their class.

2. In both 8th and 12th grade, ethnic groups differed on mathematics test scores, patterns of non-response, and metacognitive and affective variables. Whites attempted more items and outperformed African-Americans and Latinos. Note, however, that ethnicity is confounded with SES, a variable known to relate strongly to achievement. Further, in 8th grade, Whites and Latinos reported engaging in more metacognitive activity than African-Americans. African-Americans in 8th grade reported investing less effort than Whites. In 12th grade, Latinos reported less worry than Whites.

3. In 8th grade, males scored higher than females on Block 3 mathematics items but no such differences were found in 12th grade. However, in 12th grade, females omitted more items and reported worrying more than males.

4. Correlations between mathematics test score and metacognitive variables were similarly low in 12th and in 8th grade. The negative correlation between perceived worry and test score was stronger in 12th grade (r = -.52)

than in 8th grade (r = -.27) The correlation between perceived mathematics ability and test score was also much stronger in 12th grade (r = .51) than in 8th grade (r = .16).

Results: Financial Incentives Pilot Study 2: (12th-grade subjects / 8th-grade mathematics test, N=170, Whites and Latinos only)

1. Univariate and Multivariate Analysis of Variance Results, 12th Grade/8th Grade Test

1.A. Treatment Effects

1.A.1. Treatment effects on mathematics achievement. There were no treatment effects on mathematics test scores in this pilot study.

1.A.2. Treatment effects on non-response. There were no treatment effects on non-response in this pilot study.

1.A.3. Treatment effects on metacognitive and affective variables. Although intercorrelations among the four metacognitive variables ranged from .65 to .73, theory and previous research has established these variables as separate constructs which may be differentially affected by treatments; therefore, the metacognitive variables were treated as four dependent variables in a multivariate analysis of variance. MANOVA results indicated a significant difference among treatment groups on the combined metacognitive variables, multivariate $F(12, 384) = 3.2$, $p < .001$. There was a treatment effect only on one of the individual metacognitive variables, self-checking, $F(1, 154) = 3.8$, $p = .01$ (see Table 35 below and Table A13 in Appendix C). Students who were offered $1 per correct item above a minimum of eight correct reported doing more self-checking than students who were offered no reward for performance.

Table 35

Financial Incentives Pilot Study 2, Grade 12: Descriptive Statistics for Perceived Self-checking by Treatment (N=170)

Treatment	n	$\overline{X}$	SD
50 cents	43	2.7	.59
$1.00 after 8	43	2.8	.47
Class	39	2.6	.64
Control	45	2.4	.53
Total	170	2.6	.57

1.B. Ethnic Differences

1.B.1. Ethnic differences in mathematics achievement. Total mathematics test scores varied with ethnicity, $F(1, 154) = 8.6, p = .004$ (see Table A14 in Appendix C). Scheffe post hoc comparisons revealed that Whites (mean score = 31.1) outperformed Latinos (mean score = 27.5) as shown in Table 36 below.

1.B.2. Ethnic differences in non-response. There were no ethnic differences in non-response in this pilot study.

1.B.3. Ethnic differences in metacognitive and affective variables. There were no differences among ethnic groups on the combined metacognitive variables. Whites reported less worry, $F(1, 153) = 16.1, p < .001$, than Latinos (see Table A15 in Appendix C and Table 37 below).

Table 36

Financial Incentives Pilot Study 2, Grade 12: Descriptive Statistics for Total Mathematics Score by Ethnicity (N=170)

Ethnicity	n	$\overline{X}$	SD
White	108	31.1	7.0
Latino	62	27.6	7.1
Total	170	29.8	7.2

Table 37

Financial Incentives Pilot Study 2, Grade 12: Descriptive Statistics for Worry by Ethnicity (N=169)

Ethnicity	n	$\bar{X}$	SD
White	108	1.4	.52
Latino	61	1.8	.71
Total	169	1.5	.62

1.C. Gender Differences

1.C.1. Gender differences in mathematics achievement. There were no gender differences in mathematics test scores in this pilot study.

1.C.2. Gender differences in non-response. There were no gender differences in non-response in this pilot study.

1.C.3. Gender differences in metacognitive and affective variables. There were no gender differences in metacognitive or affective variables.

2. Correlations, 12th Grade/8th Grade Test

2.A. Overall mathematics performance was significantly and moderately correlated with worry, perceived mathematics ability, and previous mathematics grades (see Table 38 below).

Table 38

Financial Incentives Pilot Study 2, Grade 12: Correlations Between Total Mathematics Score and Metacognitive/Affective Variables (Ns indicated in parentheses)

	A	CS	P	SC	W	E	C	PMG	PMA
Math	.12	.06	.07	.01	-.51*	.04	-.32*	.42*	.41*
Total	(164)	(170)	(170)	(170)	(169)	(170)	(170)	(88)	(84)

Note. A = Awareness; CS = Cognitive strategy use; P = Planning; SC = Self-checking; W = Worry; E = Effort; C = Curiosity; PMG = Previous mathematics grades; PMA = Perceived mathematics ability.

*$p<.01$ (two-tailed).

3. Summary of Results (Financial Incentives Study 2, 12th grade, N=170)

1. Treatment did not affect test scores. However, students who were offered $1 per correct item over 8 reported more self-checking than students who were offered no reward.

2. Whites reported worrying less than Latinos. The test scores of Whites were also higher than the test scores of Latinos.

3. Overall mathematics performance was significantly correlated with worry, curiosity, perceived mathematics ability and previous mathematics grades.

Discussion and Implications of Results of Both Financial Incentives Studies for the Design of the Main Study

There were no differences between the mathematics performance of 8th- or 12th-grade students who were offered three types of financial incentives or standard NAEP test instructions. No differences were found either when data for subjects tested first were analyzed. However, since some increases in reported metacognitive activity were found, and since other studies of the effects of financial incentives on test performance have produced significant results, it was decided to include one financial incentive condition in the main study, and to increase the 50 cents per correct item to $1 per item for all items answered correctly.

In the focus group pilot study, 12th-grade students had indicated that they would also be motivated to try harder if they could obtain a letter of recommendation or certificate of accomplishment that could be included with their application for admission to college. Therefore, a "certificate" incentive condition was included for 12th-grade students in the main study, in addition to the financial incentive condition.

In all pilot studies, in schools where class periods were 45 or 50 minutes, a large number of students, particularly 8th graders, did not finish the self-assessment questionnaire due to a lack of time. Therefore, the number of items in the questionnaire was reduced for the main study.

White students attempted more test items and obtained higher test scores than African-Americans and Latinos in both 8th and 12th grade, regardless of which test instructions they received. This is consistent with ethnic differences found nationally on NAEP mathematics tests (Mullis et al., 1991), but confounding with SES must be considered a likely explanation.

In 8th grade, males scored higher than females on Block 3 mathematics items (the easier block). Both 8th-grade and 12th-grade females reported worrying more than males. Females' high level of perceived worry may have reduced the cognitive capacity available for processing test-relevant information, making it seem as if they were investing more effort to retrieve and apply their mathematics knowledge.

In terms of implications for the main study, the results of the financial incentives pilot studies indicated that:

1. the effects of a financial incentive of $1 per item should be investigated further as part of the main study;
2. gender and ethnicity should be retained as independent variables of interest; and
3. the number of items in the self-assessment questionnaire should be reduced.

B. Goal Orientation Study

The goal orientation pilot study compared the mathematics performance of another four groups of 8th- and 12th-grade students. Each subject received either the standard NAEP mathematics test instructions or one of the following three instructions which stated various goals of the test: (a) to compare each student's mathematical ability to that of other students (EGO); (b) to provide the opportunity for a personal accomplishment (TASK); or (c) to evaluate the effectiveness of their mathematics teacher (TEACHER). See complete text of instructions in Appendix D. Half of the students in each treatment group received the easier block of mathematics items prior to the more difficult block, and half receive the difficult set prior to the easier set.

Procedure (Goal Orientation Pilot Study)

Subjects and assignment to treatment groups. Two hundred and eight 8th-grade students and 249 12th-grade students from four schools in Southern California were tested. Schools were selected to provide a range of socioeconomic and ethnic backgrounds. An honorarium of $75 per class was paid to each school that participated. Table 39 shows the ethnic breakdown of the subjects.

Table 39

Goal Orientation Pilot Study: Ethnic Breakdown of Sample by Grade

Ethnic Group	8th grade	12th grade
White	102	128
African-American	71	21
Latino	13	69
Asian	19	27
Other	3	4
Total	208	249

The numbers of males and females in the sample are summarized in Table 40.

Table 40

Goal Orientation Pilot Study: Gender Breakdown of Sample by Grade

Gender	8th grade	12th grade
Male	101	116
Female	107	133
Total	208	249

For each grade level, subjects within each ethnic group were randomly assigned (across schools) to 8 treatment conditions. There were 8 treatment conditions because the order of the easy and difficult mathematics blocks was varied within each treatment. The numbers of students assigned to each condition are displayed in Table 41 (numbers in cells within each grade level are not equal because some subjects who were initially assigned to treatment conditions were absent from school on the day that the test was administered).

Since the results of an ANOVA indicate that neither the main effect of order nor its interaction with treatment was significant, for subsequent analysis purposes, the number of treatments was reduced to four, reflecting the three experimental motivation conditions and the control group. Latinos, Asians and students in the "Other" ethnic category were not included in the analysis of 8th-grade data, and African-Americans, Asians and students with "Other" ethnicities were not included in the analysis of 12th-grade data because there were very few students in those categories. This left a total of 173 8th-grade and 197 12th-grade students for whom data were analyzed. Tables 42 and 43 below show the final number of subjects in each cell of the treatment by ethnicity by gender design.

Table 41

Goal Orientation Pilot Study: Treatment Breakdown of Sample by Grade

Ethnic group	8th grade	12th grade
Ego, Easy first	23	32
Ego, Difficult first	29	30
Task, Easy first	30	33
Task, Difficult first	24	33
Teacher, Easy first	30	34
Teacher, Difficult first	24	30
Control, Easy first	28	29
Control, Difficult first	20	28
Total	208	249

Table 42

Goal Orientation Pilot Study, Grade 8: Number of Subjects Tested by Treatment, Ethnicity and Gender (N=173)

Treatment Group	Ethnicity								
	White			African-American			Total		
	M	F	All	M	F	All	M	F	All
Ego	11	15	26	10	6	16	21	21	42
Task	12	13	25	8	11	19	20	24	44
Teacher	11	16	27	8	12	20	29	28	47
Control	9	15	24	9	7	16	28	22	40
Total	43	59	102	35	36	71	78	95	173

Table 43

Goal Orientation Pilot Study, Grade 12: Number of Subjects Tested by Treatment, Ethnicity and Gender (N=197)

Treatment Group	Ethnicity								
	White			Latino			Total		
	M	F	All	M	F	All	M	F	All
Ego	13	21	34	7	9	16	20	30	50
Task	17	18	35	4	13	17	21	31	52
Teacher	17	15	32	6	13	19	23	28	51
Control	14	13	27	8	9	17	22	22	44
Total	61	67	128	25	44	69	86	111	197

Materials and administration. Each student received a booklet which contained two blocks of 8th-grade or 12th-grade mathematics released items from the 1990 NAEP mathematics test and a self-assessment questionnaire that consisted of 53 items. Fifty-one of the items represented four metacognitive variables (perceived planning, self-checking, cognitive strategy use, and awareness), as well as perceived effort, curiosity, and worry. The final two questions asked students to report their average grade in mathematics at the end of the previous semester, and to rank their mathematics ability compared to their classmates.

A standard script was developed for administration of the test booklets (see administration script in Appendix B). The booklets were administered by trained administrators during one regular class period. The length of class periods ranged from 45 to 55 minutes. Students tested in the shorter class periods were less likely to complete all items in the self-assessment questionnaire, since that was the last part of the booklet. In each school, administrations were sequenced during the school day; therefore, some classes were tested before others.

Follow-up with students. Approximately one month following data collection, test administrators went back to each school with a letter for each student. The letter contained information about the student's score on the mathematics items, and the 25th and 50th percentile scores on those items based on the 1990 NAEP data.

Results: Goal Orientation Pilot Study (8th grade, N=173, Whites and African-Americans; and 12th grade, N=197, Whites and Latinos)

The analyses conducted and presentation of results are similar to those for the financial incentives pilot studies described in the previous section.

1. Univariate and Multivariate Analysis of Variance Results, 8th Grade

1.A. Treatment Effects

1.A.1. Treatment effects on mathematics achievement. Treatment had no effect on mathematics test scores when data from the entire sample of 8th graders (N = 173) were analyzed. However, when only data for those students

in classes tested first in schools were analyzed (n = 55), total mathematics scores varied with treatment, $F(3, 51) = 3.4$, $p = .025$ (see Table A16 in Appendix C). Scheffe post hoc comparisons indicated that students who received the "Ego" test instructions (mean score = 28.1) had higher scores than students who received the standard NAEP instructions (mean score = 18.2); the means and standard deviations for all four groups are presented in Table 44 below. The groups were approximately ethnically balanced.

1.A.2. Treatment effects on non-response. There were no treatment effects on non-response variables.

1.A.3. Treatment effects on metacognitive and affective variables. There were no treatment effects on metacognitive or affective variables.

1.B. Ethnic Differences

1.B.1. Ethnic differences in mathematics achievement. Total score on the mathematics test varied with ethnicity, $F(1, 157) = 30.2$, $p < .001$ (see Table A17 in Appendix C). Scheffe post hoc comparisons indicated that Whites had higher scores (mean score = 24.1) than African-Americans (mean score = 17.3) (see means and standard deviations in Table 45 below).

Table 44

Goal Orientation Pilot Study, Grade 8: Descriptive Statistics for Total Mathematics Score by Treatment (N=55, students tested first)

Treatment	n	$\bar{X}$	SD
Ego	17	28.1	8.0
Task	13	24.7	11.2
Teacher	12	20.0	10.3
Control	13	18.2	7.8
Total	55	23.2	9.9

Table 45

Goal Orientation Pilot Study, Grade 8: Descriptive Statistics for Total Mathematics Score by Ethnicity (N=173)

Ethnicity	n	$\bar{X}$	SD
White	102	24.3	8.9
African-American	71	17.4	5.8
Total	173	21.5	8.5

1.B.2. Ethnic differences in non-response. Patterns of non-response did not vary with ethnicity.

1.B.3. Ethnic differences in metacognitive and affective variables. There were no ethnic differences in metacognitive or affective variables

1.C. Gender Differences

1.C.1. Gender differences in mathematics achievement. Test scores did not vary with gender.

1.C.2. Gender differences in non-response. Number of items not reached, $F(1, 157) = 6.4$, $p = .012$, and number of items not attempted, $F(1, 157) = 5.6$, $p = .019$ varied with gender (see Tables A18 and A19 in Appendix C). Males got less far in the test and attempted fewer items than did females (see Tables 46 and 47 below).

Table 46

Goal Orientation Pilot Study, Grade 8: Descriptive Statistics for Number of Mathematics Items Not Reached by Gender (N=173)

Gender	n	$\bar{X}$	SD
Male	78	3.3	4.3
Female	95	2.2	3.6
Total	173	2.9	4.1

Table 47

Goal Orientation Pilot Study, Grade 8: Descriptive Statistics for Number of Mathematics Items Not Attempted by Gender (N=173)

Gender	n	$\bar{X}$	SD
Male	78	4.4	4.8
Female	95	3.1	4.0
Total	173	3.7	4.4

1.C.3. Gender differences in metacognitive and affective variables. There were no gender differences in metacognitive or affective variables.

2. Univariate and Multivariate Analysis of Variance Results, 12th Grade

2.A. Treatment Effects

2.A.1. Treatment effects on mathematics achievement. Treatment interacted with gender in its effect on scores on Block 3 mathematics items, $F(3, 181) = 2.9$, $p = .034$ (see Table A20 in Appendix C). Analysis of simple main effects revealed that for females, scores on Block 3 varied with treatment, $F(3, 181) = 5.59$, $p = .001$). Scheffe post hoc comparisons indicated that female students who received the ego-orienting test instructions and those who received the standard NAEP instructions both outperformed females who received the teacher-orienting instructions (see Table 48 and Figure 1 below).

Table 48

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Mathematics Block 3 by Treatment and Gender (N=197)

	Male			Female			Total		
Treatment	n	$\bar{X}$	SD	n	$\bar{X}$	SD	n	$\bar{X}$	SD
Ego	20	16.2	3.9	30	15.4	5.5	50	15.8	4.9
Task	21	16.8	2.7	31	14.1	4.6	52	15.2	4.2
Teacher	23	16.2	4.2	28	12.2	4.9	51	14.0	5.0
Control	22	15.4	4.0	22	16.3	3.9	44	15.8	4.0
Total	86	16.2	3.8	111	14.4	5.0	197	15.2	4.6

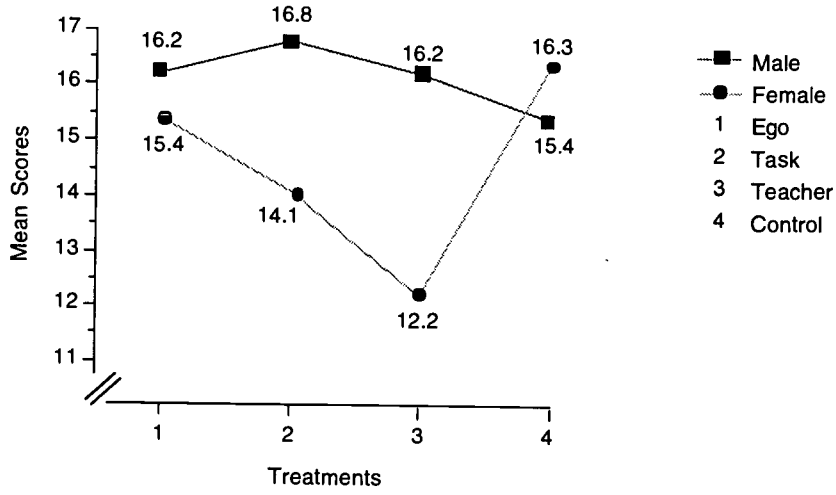

Figure 1. Mathematics Block 3 by Treatment and Gender (N=197).

2.A.2. Treatment effects on non-response. Non-response did not vary with treatment.

2.A.3. Treatment effects on metacognitive and affective variables. MANOVA revealed that treatment interacted with ethnicity in its effect on the

combined metacognitive variables, multivariate $F(12,466) = 1.88$, $p = .034$. Follow-up ANOVAs were not significant. Analysis of simple main effects and Scheffe post hoc comparisons on a variable that represented a linear combination of the four metacognitive variables revealed that Latinos who received the "teacher" instructions reported more metacognitive activity (mean = .49) than Latinos who received the standard NAEP test instructions (mean = -.6). The raw discriminant function coefficients used to form the linear combination were .95 (perceived cognitive strategy use), 2.28 (perceived self-checking), -2.63 (perceived planning) and -.51 (awareness).

2.B. Ethnic Differences

2.B.1. Ethnic differences in mathematics achievement. Test scores varied with ethnicity, $F(1, 181) = 37.4$, $p < .001$ (see Table A21 in Appendix C). The mean score for Whites (28.8) was higher than the mean score for Latinos (20.8), as shown in Table 49 below.

2.B.2. Ethnic differences in non-response. Ethnic groups differed in number of items omitted, $F(1, 181) = 8.9$, $p = .003$, number of items not reached, $F(1, 181) = 28.9$, $p < .001$, and number of items not attempted, $F(1, 181) = 34.6$, $p < .001$ (see Tables A22, A23, and A24 in Appendix C). Latinos omitted more, reached less, and attempted fewer items than did Whites (see Tables 50, 51, and 52 below).

Table 49

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Total Mathematics Score by Ethnicity (N=197)

Ethnicity	n	$\bar{X}$	SD
White	128	28.8	7.9
Latino	69	20.8	7.8
Total	197	26.0	8.7

Table 50

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Number of Mathematics Items Omitted by Ethnicity (N=197)

Ethnicity	n	$\bar{X}$	SD
White	128	.8	1.3
Latino	69	1.5	2.0
Total	197	1.0	1.6

Table 51

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Number of Mathematics Items Not Reached by Ethnicity (N=197)

Ethnicity	n	$\bar{X}$	SD
White	128	1.6	2.4
Latino	69	4.8	4.6
Total	197	2.7	3.7

Table 52

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Number of Mathematics Items Not Attempted by Ethnicity (N=197)

Ethnicity	n	$\bar{X}$	SD
White	128	2.3	2.9
Latino	69	6.3	5.1
Total	197	3.7	4.3

2.B.3. Ethnic differences in metacognitive and affective variables. Although intercorrelations among the four metacognitive variables ranged from .72 to .82, theory and previous research has established these variables as separate constructs which may be differentially affected by treatments; therefore, the metacognitive variables were treated as four dependent variables in a multivariate analysis of variance. Ethnic groups differed on the combined metacognitive variables, multivariate $F(4, 176) = 4.2$, $p = .003$. Follow-up ANOVAs indicated an ethnic difference only on perceived planning, $F(1, 179 = 5.7$, $p = .018$. Latinos reported doing more planning than Whites (see Table A25 in Appendix C and Table 53 below).

There were also ethnic differences in reported curiosity, $F(1, 179) = 17.7$, $p < .001$, and perceived worry, $F(1, 179) = 29.9$, $p < .001$ (see Tables A26 and A27 in Appendix C). Latinos reported that they were more curious and more worried than Whites (see Tables 54 and 55 below).

There were ethnic differences in perceived mathematics ability, $F(1, 166) = 5.3$, $p = .02$, and in reported previous mathematics grades, $F(1, 163) = 7.4$, $p < .007$ (see Tables A28 and A29 in Appendix C). Whites reported higher perceived mathematics ability and higher previous mathematics grades than Latinos (see Tables 56 and 57 below).

Table 53

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Planning by Ethnicity (N=195)

Ethnicity	n	$\bar{X}$	SD
White	127	2.5	.64
Latino	68	2.7	.57
Total	195	2.6	.62

Table 54

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Worry by Ethnicity (N=195)

Ethnicity	n	$\bar{X}$	SD
White	127	1.5	.52
Latino	68	2.1	.73
Total	195	1.7	.66

Table 55

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Curiosity by Ethnicity (N=195)

Ethnicity	n	$\bar{X}$	SD
White	127	2.0	.78
Latino	68	2.5	.71
Total	195	2.1	.79

Table 56

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Perceived Mathematics Ability by Ethnicity (N=182)

Ethnicity	n	$\bar{X}$	SD
White	122	3.4	.99
Latino	60	2.9	.90
Total	182	3.2	.98

Table 57

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Previous Mathematics Grades by Ethnicity (N=179)

Ethnicity	n	$\bar{X}$	SD
White	121	3.7	1.0
Latino	58	3.2	1.1
Total	179	3.5	1.1

2.C. Gender Differences

2.C.1. Gender differences in mathematics achievement. Males (mean score = 27.8) obtained higher test scores than females (mean score = 24.6), $F(1, 181), = 5.5, p = .020$ (see Table A21 in Appendix C and Table 58 below).

2.C.2. Gender differences in non-response. There were gender differences in number of items not reached, $F(1, 181) = 8.1, p = .005$, and number of items not attempted, $F(1, 181) = 8.9, p = .003$ (see Tables A23 and A24 in Appendix C). Males reached more items and attempted more items than females (see Tables 59 and 60 below).

Table 58

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Total Mathematics Score by Gender (N=197)

Gender	n	$\bar{X}$	SD
Male	86	27.8	7.6
Female	111	24.6	9.3
Total	197	26.0	8.7

Table 59

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Number of Mathematics Items Not Reached by Gender (N=197)

Gender	n	$\bar{X}$	SD
Male	86	2.0	2.8
Female	111	3.3	4.1
Total	197	2.7	3.7

Table 60

Goal Orientation Pilot Study, Grade 12: Descriptive Statistics for Number of Mathematics Items Not Attempted by Gender (N=197)

Gender	n	$\bar{X}$	SD
Male	86	2.8	3.6
Female	111	4.8	4.6
Total	197	3.7	4.3

2.C.3. Gender differences in metacognitive and affective variables. There were no gender differences on metacognitive or affective variables.

3. Correlations, 8th Grade

Overall mathematics performance was significantly correlated with worry and previous mathematics grades (see Table 61 below).

Table 61

Goal Orientation Pilot Study, Grade 8: Correlations Between Total Mathematics Score and Metacognitive/Affective Variables (Ns indicated in parentheses)

	A	CS	P	SC	W	E	C	PMG	PMA
Math	-.03	.10	.04	.11	-.45*	.04	-.04	.44*	.21*
Total	(57)	(173)	(173)	(173)	(172)	(173)	(173)	(56)	(56)

Note. A = Awareness; CS = Cognitive strategy use; P = Planning; SC = Self-checking; W = Worry; E = Effort; C = Curiosity; PMG = Previous mathematics grades; PMA = Perceived mathematics ability.

*$p<.01$ (two-tailed).

4. Correlations, 12th Grade

4.A. As shown in Table 62 below, all correlations except those between mathematics test score and planning, self-checking, and curiosity were significant (range = .17 to .57). The highest correlations were with perceived mathematics ability (r = .57), previous mathematics grades (.49) and worry (-.54). All correlations, expect with planning, self-checking and curiosity, were higher in 12th grade than in 8th grade, particularly the correlation between perceived mathematics ability and test score.

Table 62

Goal Orientation Pilot Study, Grade 12: Correlations Between Total Mathematics Score and Metacognitive/Affective Variables (Ns indicated in parentheses)

	A	CS	P	SC	W	E	C	PMG	PMA
Math	.21**	.28**	.07	.07	-.54**	.17*	-.05	.49**	.57**
Total	(195)	(195)	(195)	(195)	(195)	(195)	(195)	(179)	(182)

Note. A = Awareness; CS = Cognitive strategy use; P = Planning; SC = Self-checking; W = Worry; E = Effort; C = Curiosity; PMG = Previous mathematics grades; PMA = Perceived mathematics ability.

*$p<.05$. **$p<.01$ (two-tailed).

5. Summary of Results (Goal Orientation Study: 8th Grade, N=173; and 12th Grade, N=197)

1. In 8th grade, for the subsample of students tested first (n = 55), students who received the ego-orienting test instructions obtained higher test scores than students who received the standard NAEP test instructions. In 12th grade, females who received standard NAEP instructions or ego-orienting instructions obtained higher scores on Block 3 of the mathematics test than did female students who received the teacher-orienting instructions.

2. There were ethnic differences in total mathematics score and in metacognitive activity in both 8th and 12th grade; in 8th grade, Whites scored higher than African-Americans; in 12th grade, Whites attempted more items, reported worrying less and engaging in less planning, reported higher perceived mathematics ability and higher previous mathematics grades, and obtained higher test scores than Latinos.

3. In 8th grade, males attempted less items than did females but males' test scores were not lower than those of females. In 12th grade, males attempted more items and obtained higher total mathematics scores than females.

4. Correlations between test score and perceived planning, self-checking, curiosity, worry and mathematics ability were generally stronger in 12th grade than in 8th grade, particularly the correlation between perceived mathematics ability and test score. Correlations between test score and planning, self-checking, and curiosity were almost zero in both grades.

Discussion and Implications of Results of Goal Orientation Pilot Study for the Design of the Main Study

For students tested first in 8th grade, test instructions encouraging a competitive (ego) goal orientation produced higher test scores than standard NAEP instructions, although students receiving the ego-orientation instructions did not report investing greater effort. This result is not consistent with studies that indicate that a task-involved goal orientation introduced during instruction leads to superior performance than ego-involved goal orientation (Graham & Golan, 1991). In light of this inconsistency, it was decided that both an ego-orienting condition and a task-orienting condition

would be included in the main study. The results of this pilot study indicate that the process by which goal orientations influence performance when introduced at the time of test taking may differ from the process by which they influence performance when introduced during learning and instructional activities.

Test scores of female students varied with treatment in 12th grade. The results are unusual in that, although females who received the ego-orienting test instructions scored higher than females who received the teacher-orienting instructions, the scores of females who received the standard NAEP instructions were also higher than those of females who received the teacher-orienting instructions. It may be that the teacher-orienting instructions had a negative impact on motivation; if the goal of the test was perceived as an evaluation of the teacher rather than of the student, then the importance of the test may have been reduced, and reduced more for females than for males. This is inconsistent with the results of Brown and Walberg's (1993) study.

As in the financial incentives pilot studies, Whites obtained higher test scores than African-Americans (in 8th grade) and Latinos (in 12th grade). As in the financial incentives pilot studies, students in the goal orientation pilot study, particularly 8th-grade students, did not complete all of the self-assessment questionnaire. Therefore, the number of items in the questionnaire was reduced for the main study.

III. MAIN STUDY

The main study compared the mathematics performance of four groups of 8th-grade and five groups of 12th-grade students. At the 8th-grade level, each subject received one of three different motivational test instructions ($1 per item financial incentive, EGO goal orientation instructions; and TASK goal orientation instructions) or the standard NAEP instructions. A fifth incentive treatment was added at the 12th-grade level: a CERTIFICATE was offered to any subject who scored in the top 10% of his or her class (see test instructions in Appendix D). Since order of presentation of easy or difficult blocks of mathematics test items did not affect performance in the pilot studies, in the main study, all subjects received the easier set of mathematics items (Block 3) before the more difficult set (Block 7).

Procedure

Subjects and Assignment to Treatment Groups

Seven hundred and forty-nine 8th-grade students and 719 12th-grade students from eighteen schools in Southern California were tested. Schools were selected to provide a range of socioeconomic and ethnic backgrounds. An honorarium of $75 per class was paid to each school that participated. The ethnic and gender breakdown of the sample are shown in Tables 63 and 64 below.

Within each school, ethnic group and gender, eighth-grade subjects within each ethnic group and gender were randomly assigned to 4 treatment conditions; 12th-grade subjects were randomly assigned (within school, ethnic group and gender) to 5 treatment conditions. The numbers of students

Table 63

Main Study: Ethnic Breakdown of Sample by Grade

Ethnic group	8th grade	12th grade
White	157	169
African-American	186	183
Latino	258	238
Asian	148	129
Total	749	719

Table 64

Main Study: Gender Breakdown of Sample by Grade

Gender	8th grade	12th grade
Male	378	334
Female	371	385
Total	749	719

assigned to each condition are displayed in Table 65 (numbers in cells within each grade level are not equal because some subjects who were initially assigned to treatment conditions were absent from school on the day that the test was administered). Tables 66 and 67 below show the number of subjects in each cell of the treatment by ethnicity by gender design for 8th and 12th grade.

Materials and Administration

Each student received a booklet which contained two blocks of mathematics released items from the 1990 NAEP mathematics test and a self-assessment questionnaire that consisted of 35 items for 12th graders and 25 items for 8th graders. In the 12th-grade questionnaire, all but two of the items represented 4 metacognitive variables (perceived planning, self-checking, cognitive strategy use, and awareness), as well as self-reported effort and worry. In the 8th-grade questionnaire, all but two of the items represented 2 metacognitive variables (perceived self-checking, and cognitive strategy use), as well as self-reported effort and worry. The final two questions on both the 8th- and 12th-grade questionnaires asked students to confirm which test instructions they received and to rank their mathematics ability compared to their classmates. The second last question served as a "manipulation check," a means of verifying that the treatments were interpreted as intended.

Table 65

Main Study: Treatment Breakdown of Sample

Treatment condition	8th grade	12th grade
$1 per item	183	138
Ego orientation	196	141
Task orientation	199	144
Control	174	158
Certificate	—	138
Total	749	719

Table 66

Number of 8th-Grade Subjects Tested by Treatment, Ethnicity and Gender (N=749)

Treatment Group	Ethnicity														
	White			African-American			Latino			Asian			Total		
	M	F	All	M	F	All	M	F	All	M	F	All	M	F	All
$1.00	20	17	37	25	23	48	28	31	59	24	15	39	97	86	183
Ego	20	23	43	25	19	44	35	36	71	23	15	38	103	93	196
Teacher	19	17	36	25	26	51	32	40	72	20	20	40	96	103	199
Control	14	27	41	24	19	43	28	28	56	16	15	31	82	89	171
Total	73	84	157	99	87	186	123	135	258	83	65	148	378	371	749

Table 67

Number of 12th-Grade Subjects Tested by Treatment, Ethnicity and Gender (N=719)

Treatment Group	Ethnicity														
	White			African-American			Latino			Asian			Total		
	M	F	All	M	F	All	M	F	All	M	F	All	M	F	All
$1.00	10	20	34	19	18	37	20	21	41	13	13	26	62	76	138
Ego	15	18	33	17	18	35	22	28	50	13	10	23	67	74	141
Task	12	14	26	20	18	38	20	30	50	14	16	30	66	78	144
Certificate	18	18	36	18	17	35	20	23	43	12	12	24	68	70	138
Control	16	24	40	18	20	38	24	30	54	13	13	26	71	87	158
Total	71	98	169	92	91	183	106	132	238	65	64	129	334	385	719

Some modifications were made to the administration script that was used in the pilot studies. However, the script still included the standard NAEP script for the mathematics test. The booklets were administered by trained administrators during one regular class period. A trial administration with an 8th-grade class indicated that the test and questionnaire could be administered in 45 minutes. The length of class periods in the schools where testing took place ranged from 45 to 60 minutes. The mathematics blocks were timed tests of 15 minutes each. In each school, all test administration occurred simultaneously in order to prevent the "diffusion of treatment" problem that had been noted in the pilot studies.

It should be noted that collection of data for the main study began the week after the uprising in Los Angeles and that testing took place during the last month of the school year (May/June, 1992).

Scoring of Open-ended Items

As in the pilot studies, in the 8th-grade mathematics test there were five, and in the 12th-grade test there were eight, open-ended items. These were scored by the same three raters who scored open-ended items in the pilot studies, according to the NAEP scoring system. For the main study, interrater agreement for the 8th-grade items ranged from 98% to 100%, and for the 12th-grade items, from 97% to 100%.

Follow-up With Students

In September 1992, a letter was mailed to each subject who participated in the main study. The letter contained information about the student's score on the mathematics items, the 25th and 50th percentile scores on those items based on the 1990 NAEP data, and a check. All students except those who were offered $1 per item correct received a check for $5. Students who were promised $1 per item correct received a check for the amount of their total mathematics score. Any student in the $1-per-item condition who scored less than 5 received $5; thus, $5 was the smallest amount of money given to any student who participated in the main study. Students in the "certificate" treatment group received $5 plus a certificate if their score was in the top 10% of their class.

Analyses Conducted on Data From Main Study

Analyses similar to the first three conducted for the pilot studies were conducted (univariate analysis of variance, multivariate analysis of variance, and correlations). The main study analyses differed from the pilot studies' analyses in the following ways:

1. perceived curiosity and previous mathematics grades were not measured in 12th grade;
2. perceived curiosity, planning, awareness, and previous mathematics grades were not measured in 8th grade;
3. all analyses were conducted and are reported below for two different samples at each grade level:
   a. full sample (8th grade: n = 749; 12th grade: n = 719)
   b. subsample who remembered which test instructions they received (8th grade: n = 444; 12th grade: n = 473). This subsample was selected because it represented students for whom we have evidence that they understood the test instructions as intended.

Presentation of Main Study Results

Results of univariate and multivariate analyses of variance are presented first in the following order: full sample, 8th grade (N=749); subsample, 8th grade (N=444) if different from full sample results; full sample, 12th grade (N=719); subsample, 8th grade (N=473) if different from full sample. Correlational analyses are presented only for the full samples of 8th grade and 12th grade students. Finally, the results are summarized and discussed.

1. Univariate and Multivariate Analysis of Variance Results, 8th Grade

1.A Full Sample, 8th Grade (N=749)

1.A.1 Treatment Effects

1.A.1.a. Treatment effects on mathematics achievement. When data for the total sample of 8th-grade students (N = 749) were analyzed, a treatment effect on score on easy mathematics test items was found, $F(3, 717) = 2.7$, $p = .043$ (see Table A30 in Appendix C). Scheffe post hoc comparisons revealed

that students who were promised $1 for every item they answered correctly scored higher (means score on easy items = 7.8) than students who were given either task-oriented instructions or standard NAEP instructions (mean score = 7.5), as shown in Table 68 below).

1.A.1.b. Treatment effects on non-response. There was no treatment effect on non-response.

1.A.1.c. Treatment effects on metacognitive and affective variables. Treatment did not affect the combined metacognitive variables. The treatment groups did differ in reported effort, $F(3, 713) = 3.22$, $p = .022$ (see Table A31 in Appendix C), but the difference between the mean score on the effort scale of the group who were offered $1 per item (mean = 3.53) was not judged to be significantly higher than the means of the other groups (see Table 69 below) when Scheffe post hoc multiple comparisons were conducted.

1.A.2. Ethnic Differences

1.A.2.a. Ethnic differences in mathematics achievement. For all students tested (N = 749), mathematics test scores differed by ethnicity, $F(3, 717) = 50$, $p < .001$ (see Table A32 in Appendix C). Scheffe post hoc comparisons indicated that Asian students (mean score = 29.2) scored higher than all three

Table 68

Main Study, Grade 8: Descriptive Statistics for Easy Mathematics Items by Treatment (N=749)

Treatment	n	$\bar{X}$	SD
$1.00	183	7.8	1.2
Ego	196	7.7	1.3
Task	199	7.5	1.6
Control	171	7.5	1.5
Total	749	7.6	1.4

Table 69

Main Study, Grade 8: Descriptive Statistics for Effort by Treatment (N=745)

Treatment	n	$\bar{X}$	SD
$1.00	183	3.53	.56
Ego	196	3.36	.65
Task	197	3.36	.63
Control	169	3.40	.64
Total	745	3.41	.63

other ethnic groups as shown in Table 70 below (White mean score = 25.9; African-American mean score = 22.2; Latino mean score = 20.4). In addition, White students' scores were significantly higher than Latinos' and African-Americans'.

1.A.2.b. Ethnic differences in non-response. There were no ethnic differences in non-response.

1.A.2.c. Ethnic differences in metacognitive and affective variables. Ethnic groups did not differ on the combined metacognitive variables, multivariate $F(6, 1422) = 1.9$, $p = .07$. Ethnic groups differed on perceived mathematics ability, $F(3, 602) = 8.4$, $p < .001$, perceived effort, $F(3, 713) = 2.9$, $p = .033$, and perceived worry, $F(3, 713) = 10.7$, $p < .001$ (see Tables A31, A33 and

Table 70

Main Study, Grade 8: Descriptive Statistics for Total Math Score by Ethnicity (N=749)

Ethnicity	n	$\bar{X}$	SD
White	157	25.9	7.3
African-American	186	22.2	8.2
Latino	258	20.4	7.1
Asian	148	29.2	7.6
Total	749	23.7	8.2

A34 in Appendix C). Latinos reported worrying more than all three other ethnic groups and had lower perceptions of their mathematics ability than either Asians or African Americans (see Tables 71 and 72 below). Scheffe post hoc comparisons revealed no significant differences among ethnic groups on perceived effort.

1.A.3. Gender Differences

1.A.3.a. Gender differences in mathematics achievement. There were no gender differences in mathematics achievement.

Table 71

Main Study, Grade 8: Descriptive Statistics for Perceived Mathematics Ability by Ethnicity (N=634)

Ethnicity	n	$\bar{X}$	SD
White	136	3.4	.87
African-American	151	3.4	.83
Latino	213	3.1	.86
Asian	134	3.6	.85
Total	634	3.4	.87

Table 72

Main Study, Grade 8: Descriptive Statistics for Worry by Ethnicity (N=745)

Ethnicity	n	$\bar{X}$	SD
White	156	1.6	.62
African-American	186	1.7	.63
Latino	256	2.0	.66
Asian	147	1.7	.60
Total	745	1.8	.64

1.A.3.b. Gender differences in non-response. In the total sample (N = 749) there was a gender difference in number of items not reached, $F(1, 717) = 4.5$, $p = .033$ (see Table A35 in Appendix C). Females got further in the test than did males (see Table 73 below).

1.A.3.c. Gender differences in metacognitive and affective variables. Males and females differed on the combined metacognitive variables, multivariate $F(2, 711) = 3.24$, $p = .040$. Follow-up univariate F tests revealed that females reported doing more self-checking than makes (see Table A36 in Appendix C and Table 74 below). Females also reported investing more effort than did males (see Table A31 in Appendix C and Table 75 below).

Table 73

Main Study, Grade 8: Descriptive Statistics for Number of Mathematics Items Not Reached by Gender (N=749)

Gender	n	$\bar{X}$	SD
Male	378	.9	2.8
Female	371	.5	1.4
Total	749	.7	2.2

Table 74

Main Study, Grade 8: Descriptive Statistics for Self-checking by Gender (N=745)

Gender	n	$\bar{X}$	SD
Male	375	2.66	.64
Female	370	2.74	.59
Total	745	2.70	.62

Table 75

Main Study, Grade 8: Descriptive Statistics for Effort by Gender (N=745)

Gender	n	$\bar{X}$	SD
Male	375	3.3	.65
Female	370	3.5	.58
Total	745	3.4	.62

1.B. Subsample, 8th Grade (N=444)

1.B.1. Treatment Effects

1.B.1.a. Treatment effects on mathematics achievement. When data for subjects who correctly identified the test instructions they received were analyzed (N = 444), the effect of treatment on total mathematics test score was significant, $F(3, 412) = 3.0$, $p = .029$ (see Table A37 in Appendix C). Students who were offered $1 for each item they answered correctly scored higher (mean score = 28.5) than students who received the standard NAEP test instructions (mean score = 25.2), as shown in Table 76 below. This difference was reflected in scores on easy, moderately difficult, and open-ended items, but not on difficult items. The difference in mean test score translates into an

Table 76

Main Study, Grade 8: Descriptive Statistics for Total Mathematics Score by Treatment (N=444)

Treatment	n	$\bar{X}$	SD
$1.00	95	28.5	7.6
Ego	124	26.0	7.9
Task	108	26.5	7.1
Control	117	25.2	8.2
Total	443	26.5	7.8

effect size of .41. In the subsample, the treatment groups also differed inreported effort, $F(3, 411) = 3.7$, $p = .012$ (see Table A38 in Appendix C). Scheffe post hoc multiple comparisons revealed that students who were offered $1 for every item they answered correctly reported investing more effort than students who got either the task-oriented or standard NAEP test instructions (see Table 77 below).

1.B.1.b. Treatment effects on non-response. There was no treatment effect on non-response.

1.B.1.c. Treatment effects on metacognitive and affective variables. Treatment effects on metacognitive and affective variables in the subsample were similar to those found in the full sample.

1.B.2. Ethnic Differences

Ethnic differences observed in the 8th-grade subsample were similar to the ethnic differences found in the full sample.

1.B.3. Gender Differences

1.B.3.a. Gender differences in mathematics achievement. As was the case for the full sample, there were no gender differences in mathematics achievement in the subsample.

1.B.3.b. Gender differences in non-response. Gender differences in non-response for the subsample were similar to the differences reported for the full sample.

Table 77

Main Study, Grade 8: Descriptive Statistics for Effort by Treatment (N=443)

Treatment	n	$\bar{X}$	SD
$1.00	95	3.6	.42
Ego	124	3.5	.61
Task	108	3.4	.56
Control	116	3.4	.60
Total	443	3.5	.56

1.B.3.c. Gender differences in metacognitive and affective variables. Unlike in the full sample, males and females did not differ on the combined metacognitive variables in the subsample. However, as was the case in the full sample, females reported investing more effort than males.

2. Univariate and Multivariate Analysis of Variance Results, 12th Grade

2.A. Full Sample, 12th Grade (N=719)

2.A.1. Treatment Effects

2.A.1.a. Treatment effects on mathematics achievement. There were no treatment effects on mathematics performance.

2.A.1.b. Treatment effects on non-response. There were no treatment effects on non-response.

2.A.1.c. Treatment effects on metacognitive and affective variables. Ratings on combined metacognitive variables varied with treatment, multivariate $F(16, 2051) = 1.8$, $p = .022$. Post hoc univariate F tests revealed no differences. However, comparison of mean scores of the treatment groups on a variable that was a linear combination of all four metacognitive variables revealed that students in the group who were offered $1 per correct item engaged in more metacognitive activity (mean = 2.38) than students who received the standard NAEP test instructions (mean = 2.0). The raw discriminant function coefficients used to form the linear combination were 1.3 (perceived self-checking), .77 (perceived cognitive strategy use), -1.82 (perceived planning), and .72 (perceived awareness).

2.A.2. Ethnic Differences

2.A.2.a. Ethnic differences in mathematics achievement. Scores on the mathematics test varied with ethnicity, $F(3, 679) = 80.7$, $p < .001$ (see Table A39 in Appendix C). Scheffe post hoc comparisons revealed that Whites (mean score = 28.8) and Asians (mean score = 30.5) outperformed African-Americans (mean score = 19.7) and Latinos (mean score = 21.6), as shown in Table 78 below.

Table 78

Main Study, Grade 12: Descriptive Statistics for Total Mathematics Score by Ethnicity (N=719)

Ethnicity	n	$\bar{X}$	SD
White	169	28.8	8.0
African-American	183	19.7	7.6
Latino	238	21.6	7.5
Asian	129	30.5	7.4
Total	719	24.4	8.8

2.A.2.b. Ethnic differences in non-response. There were ethnic differences on all three non-response variables: number of items omitted, $F(3, 679) = 3.8$, $p = .01$, number of items not reached, $F(3, 679) = 9.9$, $p < .001$, and number of items not attempted, $F(3, 679) = 11.1$, $p < .001$ (see Tables A40, A41, and A42 in Appendix C). African-Americans omitted more items, did not get as far in the test, and consequently attempted fewer items than either Asians or Whites. Latinos did not reach as many items and attempted fewer items than either Asians or Whites (see Tables 79, 80, and 81 below).

Table 79

Main Study, Grade 12: Descriptive Statistics for Number of Mathematics Items Omitted by Ethnicity (N=719)

Ethnicity	n	$\bar{X}$	SD
White	169	.7	1.0
African-American	183	1.1	1.8
Latino	238	.8	1.2
Asian	129	.6	1.2
Total	719	.8	1.4

Table 80

Main Study, Grade 12: Descriptive Statistics for Number of Mathematics Items Not Reached by Ethnicity (N=719)

Ethnicity	n	$\bar{X}$	SD
White	169	1.4	2.4
African-American	183	2.8	4.1
Latino	238	2.4	2.7
Asian	129	1.4	1.8
Total	719	2.1	3.0

Table 81

Main Study, Grade 12: Descriptive Statistics for Number of Mathematics Items Not Attempted by Ethnicity (N=719)

Ethnicity	n	$\bar{X}$	SD
White	169	2.1	2.6
African-American	183	3.9	5.0
Latino	238	3.2	3.0
Asian	129	2.0	2.2
Total	719	2.9	3.5

2.A.2.c. Ethnic differences in metacognitive and affective variables. Ethnic groups differed on the combined metacognitive variables, multivariate $F(12, 1776) = 1.90$, $p = .030$. Follow-up univariate F tests revealed that ethnic groups differed on perceived self-checking only (see Table A43 in Appendix C and Table 82 below); however, Scheffe post hoc comparisons revealed no significant differences. Comparison of mean scores on a variable that was a linear combination of all four metacognitive variables revealed that Asians (mean = 2.7) and Latinos (mean = 2.6) reported more metacognitive activity than African-Americans (mean = 2.3). The raw discriminant function coefficients used to form the linear combination were 1.9 (perceived self-checking), -.55 (perceived cognitive strategy use), -1.17 (perceived awareness),

Table 82

Main Study, Grade 12: Descriptive Statistics for Self-checking by Ethnicity (N=715)

Ethnicity	n	$\bar{X}$	SD
White	169	2.6	.65
African-American	181	2.4	.63
Latino	237	2.6	.64
Asian	128	2.6	.64
Total	715	2.6	.64

and .90 (perceived planning). Perceived mathematics ability, $F(3,669) = 9.3$, $p < .001$, worry, $F(3, 675) = 2.4$, $p = .022$, and effort, $F(3,675) = 8.9$, $p < .001$, varied with ethnicity (see Tables A44, A45, and A46 in Appendix C). Latinos and Asians reported more worry than Whites, and Latinos reported more worry than African-Americans; Latinos, Whites and Asians reported investing more effort than African-Americans; Asians had higher perceived mathematics ability than Latinos and African-Americans (see Tables 83, 84, and 85 below).

Table 83

Main Study, Grade 12: Descriptive Statistics for Worry by Ethnicity (N=715)

Ethnicity	n	$\bar{X}$	SD
White	169	1.5	.59
African-American	181	1.7	.52
Latino	237	1.9	.65
Asian	128	1.8	.69
Total	715	1.7	.63

Table 84

Main Study, Grade 12: Descriptive Statistics for Effort by Ethnicity (N=719)

Ethnicity	n	$\bar{X}$	SD
White	169	15.4	3.5
African-American	183	13.7	4.1
Latino	238	15.4	3.7
Asian	129	15.9	3.7
Total	719	15.1	3.8

Table 85

Main Study, Grade 12: Descriptive Statistics for Perceived Mathematics Ability by Ethnicity (N=670)

Ethnicity	n	$\bar{X}$	SD
White	163	3.3	.89
African-American	164	3.1	.71
Latino	219	3.1	.75
Asian	124	3.5	.86
Total	670	3.2	.81

2.A.3. Gender Differences

2.A.3.a. Gender differences in mathematics achievement. Males (mean score = 25.5) obtained higher test scores than females (mean score = 23.5), $F(1, 679) = 12.4$, $p < .001$ (see Table A37 in Appendix C and Table 86 below).

2.A.3.b. Gender differences in non-response. There were no gender differences on non-response variables.

2.A.3.c. Gender differences in metacognitive and affective variables. Males and females differed on the combined metacognitive variables, multivariate $F(4, 671) = 5.37$, $p < .001$. Post hoc univariate F tests revealed a significant difference on perceived self-checking (see Table A43 in Appendix C

and Table 87 below). Females reported doing more self-checking than males. Perceived effort, $F(1, 675) = 7.7$, $p = .006$, and perceived mathematics ability, $F(1, 630) = 13.6$, $p < .001$, also varied with gender (see Tables A45 and A46 in Appendix C). Females reported investing more effort and having lower perceptions of their mathematics ability than males (see Tables 88 and 89).

Table 86

Main Study, Grade 12: Descriptive Statistics for Total Mathematics Score by Gender (N=719)

Gender	n	$\bar{X}$	SD
Male	334	25.5	9.1
Female	385	23.5	8.4
Total	719	24.4	8.8

Table 87

Main Study, Grade 12: Descriptive Statistics for Self-checking by Gender (N=715)

Gender	n	$\bar{X}$	SD
Male	331	2.5	.67
Female	384	2.6	.62
Total	715	2.6	.64

Table 88

Main Study, Grade 12: Descriptive Statistics for Effort by Gender (N=715)

Gender	n	$\bar{X}$	SD
Male	331	3.0	.74
Female	384	3.1	.69
Total	715	3.1	.72

Table 89

Main Study, Grade 12: Descriptive Statistics for Perceived Mathematics Ability by Gender (N=670)

Gender	n	$\bar{X}$	SD
Male	307	3.3	.83
Female	363	3.1	.78
Total	670	3.2	.81

2.B. Subsample, 12th Grade (N=473)

Results for the subsample did not differ from results for the full sample.

3. Correlations, 8th Grade, Full Sample (N=749)

3.1. Correlations between total mathematics score and metacognitive and affective variables. Table 90 below shows that mathematics performance was significantly correlated with all metacognitive and affective variables, the highest correlations being with worry (r = -.45) and perceived mathematics ability (r = .42). These two correlations indicate that as worry increased, test performance declined; as perceived mathematics ability increased, test performance also increased.

Table 90

Main Study, Grade 8: Correlations Between Total Mathematics Score and Metacognitive/Affective Variables (Ns indicated in parentheses)

	CS	SC	W	E	PMG
Math Total	.15** (745)	.17** (744)	-.45** (745)	.24** (745)	.42** (634)

Note. CS = Cognitive strategy use; SC = Self-checking; W = Worry; E = Effort; PMA = Perceived mathematics ability.

*p<.05. **p<.01 (two-tailed).

3.2. Intercorrelations among metacognitive and affective variables. Table 91 below indicates that the correlations between metacognitive variables (perceived self-checking and cognitive strategy use) and perceived effort were around .5. Perceived effort was not related to perceived worry or perceived mathematics ability. Metacognitive variables were weakly related to perceived mathematics ability but not to perceived worry. Worry was negatively correlated (-.29) with perceived mathematics ability.

4. Correlations, 12th Grade

4.A. Full Sample, 12th Grade (N=719)

4.A.1. Correlation between total mathematics score and metacognitive and affective variables. Table 92 below shows that mathematics performance was significantly correlated with all metacognitive and affective variables, the highest correlations being with worry (-.36) and perceived mathematics ability (.48). These two correlations indicate that as worry increased, test performance declined; as perceived mathematics ability increased, test performance also increased. This pattern of correlations is similar to that found in 8th grade.

Table 91

Main Study, Grade 8: Intercorrelations Among Metacognitive and Affective Variables (Ns indicated in parentheses)

	1	2	3	4	5
1. Cognitive Strategy	1.00				
2. Self-checking	.55** (744)	1.00			
3. Worry	.08* (745)	-.11** (744)	1.00		
4. Effort	.51** (745)	.54** (744)	.04 (745)	1.00	
5. Perceived Math Ability	.20** (634)	.21 (634)	-.29 (634)	.08* (634)	1.00

*p<.05 (two-tailed). **p<.01 (two-tailed).

Table 92

Main Study, Grade 12: Correlations Between Total Mathematics Score and Metacognitive/Affective Variables (Ns indicated in parentheses)

	CS	SC	W	E	P	A	PMA
Math Total	.21** (714)	.20** (715)	-.36** (715)	.22** (715)	.17** (715)	.21** (715)	.48** (670)

Note. CS = Cognitive strategy use; SC = Self-checking; W = Worry; E = Effort; P = Planning; A = Awareness; PMA = Perceived mathematics ability

*p<.05. **p<.01 (two-tailed).

4.A.2. Intercorrelations among metacognitive and affective variables. Table 93 indicates that perceived effort was highly correlated with metacognitive variables (range = .59 to .65). Perceived effort was not related to perceived worry and was only weakly related to perceived mathematics ability (r = .15). Metacognitive variables were weakly related to perceived mathematics ability, but not to perceived worry. Worry was negatively related to perceived mathematics ability (r = -.31). This pattern of correlations is similar to that found in the 8th grade.

5. Main Study: Summary of Results

5.1 Full Sample, 8th Grade (N=749)

5.1.a. Treatment effects. In 8th grade, students who were offered a financial incentive for test performance ($1 per item correct) obtained higher scores on easier test items than did students who received standard NAEP test instructions.

5.1.b. Ethnic differences. In 8th grade, Asian students scored higher than all other ethnic groups, and Whites scored higher than African-Americans and Latinos. Latinos reported lower perceived mathematics ability than Asians and African-Americans. Latinos reported worrying more than all three other ethnic groups.

Table 93

Main Study, Grade 12: Intercorrelations Among Metacognitive and Affective Variables (Ns indicated in parentheses)

	1	2	3	4	5	6	7
1. Cognitive Strategy	1.00						
2. Self-checking	.70** (714)	1.00					
3. Worry	.01 (714)	-.01 (715)	1.00				
4. Effort	.60** (714)	.64** (715)	.01 (715)	1.00			
5. Planning	.79** (714)	.67** (715)	.04 (715)	.59** (715)	1.00		
6. Awareness	.74** (714)	.69** (715)	-.02 (715)	.65** (715)	.72** (715)	1.00	
7. Perceived Math Ability	.21** (670)	.19** (670)	-.31** (670)	.15** (670)	.23** (670)	.19** (670)	1.00

*p<.05 (two-tailed). **p<.01 (two-tailed).

5.1.c. Gender differences. In 8th grade, there was no difference between the test scores of males and females. However, females got further in the test than males, and females reported more effort and more self-checking than males.

5.1.d. Correlations. Worry and perceived mathematics ability were most highly correlated with test score, the relationship between worry and test performance being negative. However, worry was not at all related to perceived mathematics ability. Effort was moderately correlated with metacognitive variables, but neither effort nor metacognitive variables were strongly correlated with worry or perceived mathematics ability.

5.2. Subsample, 8th Grade (N=444)

In 8th grade, results for the subsample of students who remembered which test instructions they received were generally similar to the results for the full sample. However, the effect of the financial incentive was stronger for

the subsample. Not only was the mean score on easier items higher for the group who received the financial incentive than for the group who received the standard NAEP test instructions, but score on moderately difficult items and on open-ended items was also higher, resulting in a higher mean overall test score. In addition, students who were offered the financial incentive of $1 per item reported investing more effort than did students in either the group who received the task-oriented instructions or the group who received the standard NAEP instructions.

5.3. Full Sample, 12th Grade (N=719)

5.3.a. Treatment effects. In 12th grade, there were no differences among the test scores of students who received different test instructions. However, the group who received the financial incentive reported more metacognitive activity than the group who got the standard NAEP test instructions.

5.3.b. Ethnic differences. In 12th grade, Asian and White students scored higher than African-Americans and Latinos. In addition, Asians and Whites attempted more items than African-Americans and Latinos, and reported more metacognitive activity than African-Americans. Asians and Latinos reported more metacognitive activity and effort than African-Americans. Whites reported investing more effort than African-Americans. Latinos reported worrying more than Whites and African-Americans; Asians reported worrying more than Whites. Asians had higher perceptions of their mathematics ability than Latinos and African Americans.

5.3.c. Gender differences. In 12th grade, males had a higher mean test score than females. However, females reported investing more effort, doing more self-checking, and having lower perceived mathematics ability than did males.

5.3.d. Correlations. The pattern of correlations in 12th grade was similar to the pattern in 8th grade. Worry and perceived mathematics ability were most highly correlated with test score, the relationship between worry and test performance being negative. Perceived effort and metacognitive variables were not related to perceived worry, and only weakly related to perceived mathematics ability. Worry was negatively related to perceived mathematics ability.

5.4 Subsample, 12th Grade (N=473)

In 12th grade, results for the subsample of students who remembered which test instructions they received were similar to the results for the full sample.

Discussion and Implications of Results of Main Study

In 8th grade, for the full sample, the financial incentive increased test scores on easier items only. In the subsample of students who remembered which test instructions they received, there was an increase in overall test score, reflecting increases in scores on easy items, moderately difficult items, and on open-ended items, but no increase in performance on difficult items. The increased performance of 8th-grade students who received $1 per item correct was accompanied by an increase in perceived effort in the subsample who remembered their test instructions. This adds support to the theory that it is through increased effort that motivation impacts performance. The increase in perceived effort was not accompanied by an increase in reported metacognition, but perceived effort was moderately to strongly correlated with the metacognitive variables. The fact that there was no increase in scores on difficult test items suggests that increased investment of effort permits greater retrieval and use of prior knowledge when one possesses relevant prior knowledge, but does not affect performance when prior knowledge is weak.

In 12th grade, only reported metacognition differed with treatment. Again, the financial incentive condition was more effective than the standard NAEP test instructions. Students who were offered $1 per correct item reported engaging in more metacognitive activity than students who received standard NAEP instructions. However, these differences in reported metacognition did not translate into differences in mathematics test scores. This suggests that, while the financial incentives led 12th-grade students to "try harder" by using more of their metacognitive skills, their mathematical knowledge may not have been sufficient to have that extra cognitive effort make a significant difference in their test scores.

Different test instructions did not have different effects on different ethnic groups. In general, in both grade levels, regardless of test instructions, Asians and Whites scored higher, reported more effort, less worry, and higher

perceptions of their mathematics ability than either Latinos or African-Americans. Because no reliable measures of social class were obtained in this study, it is not clear whether the observed ethnic differences are in fact ethnic differences or social class differences. The ethnic differences in worry and perceived mathematics ability found in this study are consistent with previous research and motivational theory which suggest that low perceptions of ability lead to higher anxiety which in turn hinders performance (Wigfield & Eccles, 1989). Worry was moderately correlated with perceived mathematics ability.

In both 8th and 12th grade, females reported investing more effort and doing more self-checking than males to achieve similar test scores in 8th grade and lower scores than males in 12th grade. These results may indicate that females either are investing more effort to compensate for a lack of prior knowledge, or have inaccurate perceptions of how much effort they are investing. In both 8th and 12th grades, perceived effort and metacognition were not as strongly correlated with test score as were perceived mathematics ability and worry. Furthermore, perceived effort and metacognition were not related to perceived mathematics ability and worry. The studies reported here attempted to affect test performance through interventions targeted at effort. Additional improvements in test performance might result from interventions that target worry and perceptions of one's ability.

In summary, the results of this study indicate that students' investment of effort and level of metacognitive activity can be manipulated by external financial rewards offered at the time of test-taking. The results also suggest that an increase in effort can be translated into an increase in test scores, at least for 8th grade students. It seems that variables that operate at the time of test taking and that influence cognitive activity, worry, effort and performance are worthy of continued research, particularly research that attempts to unravel the complex causal paths among these variables.

REFERENCES

Abedi, J. (1974). *A comparison of statistical techniques for testing for heterogeneity of variances*. Unpublished dissertation, George Peabody College for Teachers, Nashville. (University Microfilms No. 75-20-261)

Ames, C. (1992). Classroom: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*, 261-271.

Ames, C., & Archer, J. (1988). Achievement in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*, 260-267.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, *84*,191-215.

Benbow C., & Stanley, J. (1982). Sex differences in mathematical ability: Fact or artifact. *Science*, *10*, 1262-1264.

Boekarts, M. (1988). Motivated learning: Bias in appraisals. *International Journal of Educational Research*, *12*, 267-280.

Brophy, J.E. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51*, 5-32.

Brown, S. M., & Walberg, H. J. (1993). Motivational effects on mathematics test scores of elementary-school students. *Journal of Educational Research*, *86*(3), 133-136.

Butler, R. (1987). Task-involving and ego-involving properties of evaluations: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, *75*, 544-552.

Corno, L. (1986). The metacognitive control components of self-regulated learning. *Contemporary Educational Psychology, 11,* 333-346.

Corno, L., & Mandinach, E.B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, *18*, 88-108.

Cotton, J.L., & Cook, M.A. (1982). Meta-analyses and the effects of various reward systems: Some different conclusions from Johnson et al. *Psychological Bulletin*, 92, 176-183.

Covington, M.V. (1986). Anatomy of failure-induced anxiety: The role of cognitive mediators. In R. Schwarzer (Ed.), *Self-related cognitions in anxiety and motivation* (pp. 247-263). Hillsdale, NJ: Erlbaum.

Covington, M.V., & Omelich, C.L. (1979). Effort: The double-edged sword in school achievement. *Journal of Educational Psychology*, *71*, 169-182.

Deci, E.L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*, 105-115.

Dweck, C.S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Dweck, C.S. (1989). Motivation. In A. Lesgold & R. Glaser (Eds.), *Foundations for a psychology of education.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Dweck, C.S., & Elliott, E.S. (1983). Achievement motivation. In P. Mussen (Ed.), *Handbook of child psychology* (pp. 643-691). New York: Wiley.

d'Ydewalle, G. (1987). Is it still worthwhile to investigate the impact of motivation on learning? In E. de Corte, H. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction* (Vol. 1, pp. 191-200). Oxford: Pergamon.

Eccles, J.S. (1983). Expectancies, values, and academic behavior. In J.C. Spencer (Ed.), *Achievement and achievement motivation* (pp. 75-146). San Francisco, CA: Freeman.

Educational Testing Service. (1991, May). *The results of the NAEP 1991 field test for the 1992 national and trial state assessments. Draft.* Princeton, NJ: Author.

Elliott, E., & Dweck, C. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*, 5-12.

Fowler, R. L., & Clingman, J. (1977). The influence of intrinsic and extrinsic reward on the intratest performance of high-and low-scoring children. *The Psychological Record, 3*, 603-610.

Fraser, B.J., Walberg, H.J., Welch, W.W., & Hattie, J.A. (1987). Synthesis of educational productivity research. *International Journal of Educational Research, 11*, 145-252.

Garcia-Celay, I.M., & Tapia, J.A. (1992) Achievement motivation in high school: Contrasting theoretical models in the classroom. *International Journal of Educational Research, 18*, 43-57.

Graham, S., & Golan, S. (1991). Motivational influences on cognition: Task involvement, ego involvement, and depth of information processing. *Journal of Educational Psychology, 83*, 187-194.

Graham, S., & Weiner, B. (1986). From attribution theory to developmental psychology: A round-trip ticket? *Social Cognition, 4*, 152-179.

Helmke, A. (1989). Affective student characteristics and cognitive development: Problems, pitfalls, perspectives. *International Journal of Educational Research, 13*, 915-932.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*, 47-77.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60*, 549-571.

Hill, K.T., & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *Elementary School Journal, 85,* 105-126.

Iowa Tests of Basic Skills normal curve equivalent norms. (1978). Boston, MA: Houghton Mifflin.

Koretz, D., Lewis, E., Burstein, L., & Skewes-Cox, T. (1992). *Omitted and not-reached items in mathematics in the 1990 National Assessment of Educational Progress* (CRESST/RAND report). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Kukla, A. (1978). An attributional theory of choice. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 113-144). New York: Academic Press.

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

MacIver, D. (1988). Classroom environments and the stratification of pupils' ability perceptions. *Journal of Educational Psychology, 80*, 495-505.

Meece, J., Blumenfeld, P., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*, 514-523.

Morgan, M. (1984). Reward-induced decrements and increments in intrinsic motivation. *Review of Educational Research, 54*, 5-30.

Morris, L.W., Davis, M.A., & Hutchings, C.H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology, 73*(4), 541-555.

Mullis, I.V.S., Dossey, J.A., Owen, E. H., & Phillips, G.W. (1991). The state of mathematics achievement: NAEP's 1990 assessment of the nation and the trial assessment of the states (Report No. 21-ST-03). Princeton, NJ: Educational Testing Service.

Nicholls, J.G. (1983). Conceptions of ability and achievement motivation: A theory and its implications for education. In S.G. Paris, G.M. Olson, & H. W. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 211-237). Hillsdale, NJ: Erlbaum.

Nicholls, J.G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328-346.

O'Neil, H.F., Jr., & Abedi, J. (1992). Japanese children's trait and state worry and emotionality in a high-stakes testing environment. *Anxiety, Stress, and Coping, 5*(3), 253-267.

O'Neil, H.F., Jr., Baker, E.L., Jacoby, A., Ni, Y., & Wittrock, M. (1990). *Human benchmarking studies of expert systems*. Los Angeles: University of California, Center for Technology Assessment.

Pintrich, P.R., & De Groot, E.V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 33-40.

Powers, D.E. (1986). *Test anxiety and the GRE General Test* (GRE Board Professional Report No. 83-17P; ETS Research Report 86-84). Princeton, NJ: Educational Testing Service.

Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychology, 18*(1), 42-50.

Schunk, D.H. (1983). Developing children's self-efficacy and skills: The roles of social comparative information and goal setting. *Contemporary Educational Psychology, 8*, 76-86.

Schunk, D.H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist, 19*, 48-58.

Schunk, D.H. (1989). Self-efficacy and cognitive skills learning. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 13-44). San Diego, CA: Academic Press.

Schunk, D.H. (1990). Introduction to the special section on motivation and efficacy. *Journal of Educational Psychology*, 82, 3-6.

Shanker, A. (1990). How much do our kids really know? Raising the stakes on NAEP. *The New York Times,* July 29.

Sieber, J.E., O'Neil, H.F., Jr., & Tobias, S. (Eds.). (1977). *Anxiety, learning and instruction.* Hillsdale, NJ: Erlbaum.

Stipek, D.J., & Weisz, J.R. (1981). Perceived personal control and academic achievement. *Review of Educational Research, 51*, 101-137.

Swinton, S. (1991). *Differential response rates to open-ended and multiple-choice NAEP items by ethnic groups.* Princeton, NJ: Educational Testing Service.

Teidman, G.L., & McMahon, R.J. (1985). Brief Reports: Individual differences in children's response to self-and externally-administered reward. *Behavior Therapy, 16*, 516-523.

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

Uguroglu, M.E., & Walberg, H.J. (1979). Motivation and achievement: A Quantitative synthesis. *American Educational Research Journal, 16*, 375-389.

Weinberg, R.S. (1978). Relationship between extrinsic rewards and intrinsic motivation. *Psychological Report, 42*, 1255-1258.

Weiner, B.J. (1986). *An attributional theory of motivation and emotion.* New York: Springer-Verlag.

Wigfield, A., & Eccles, J.S. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist, 24*,159-183.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76*, 92-104.

Winer, B.J., Brown, D.R., & Michels, K.M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Zimmerman, B.J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology, 11*, 307-313.

Zimmerman, B.J., & Martin-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*, 51-59.

APPENDIX A: History, Revision, and Validation of the Metacognitive Skill Instrument*

* The instruments were revised under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

Overview

One of the key domain-independent variable constructs believed to be useful in measuring indirectly whether students are motivated is self-regulation. It is expected that when students are motivated, their self-regulation skills would be engaged. We define self-regulation as metacognitive skills and effort. To test if this in fact is true, a battery of metacognitive and affective measures was adapted. This battery originally consisted of 100 items, which included the following:

1. State measures of metacognition (planning, self-checking, cognitive strategies, awareness) by Harold F. O'Neil, Jr. (O'Neil, Baker, Jacoby, Ni, & Wittrock, 1990);

2. State measures of effort developed by Harold F. O'Neil, Jr. and Richard Snow;

3. State measures of worry and emotionality. The state versions of the measures were revised scales originally developed by Morris, Davis and Hutchings (1981) and modified, based on back-translations of a Japanese state worry and emotionality scale (O'Neil, Baker, & Matsuura, 1992) by O'Neil;

4. A state measure of curiosity developed by Spielberger, Peters, and Frain (1976, 1981).

This 100-item state questionnaire was administered to a group of 236 junior college students to examine its psychometric characteristics (Kosmicki, 1993). Descriptive statistics such as mean, standard deviation, measures of skewness and kurtosis, as well as frequency distributions, univariate and bivariate graphs, were obtained for each item and each subscale. A classical measure of reliability, Cronbach's Alpha, was obtained to examine internal consistency for the items in each subscale. To further evaluate the internal consistency of items within the subscales, factor analysis was applied to items in subscales. A mathematics achievement test score was used as a criterion to see if there is any relationship between the scores of this test with the subscales of the metacognitive/ affective instrument, that is, to get an estimate of concurrent validity of the instrument. Based on the descriptive statistics, internal consistency measures, and the results of factor analysis and validity studies, poor items were identified and removed, and the number of items was reduced from 100 to 70. The elimination of items was carefully done so that no

significant reduction in the reliability or validity indices of the subscales was observed. The reduced form of the state metacognitive questionnaire was administered to another group of 210 high school students (Khabiri, 1993). The same type of analyses were performed on the reduced form, and, based on the results, the state items were further reduced from 70 to 50. The pool of metacognitive items resulting from the second administration was used in the pilot phase of the experimental motivation study on 376 8th-grade and 464 12th-grade students.

The results of the pilot studies, however, suggested that the majority of 8th-grade students (and a few 12th graders) could not even complete the reduced 50-item instrument within the time constraints of administering two NAEP blocks (15 minutes each) and instructions within one class period of less than one hour. We decided to use the results of the pilot studies to see if a shorter version of the instrument were possible. The results of the statistical analyses on the pilot studies' data and NCES staff input on item sensitivity indicated that the reliability and validity of subscales could remain at an acceptable level with a minimum of five items in each subscale, but further reducing the number of items could seriously affect reliability and validity of subscales. The high correlations between subscale scores, and between subscale scores and math performance, however, suggested the possibility of shortening the instrument for 8th graders by omitting a few of the subscales. Since the number of unreached self-assessment items was much greater for 8th-graders than 12th-graders, we decided to omit the planning and awareness subscales for the 8th-grade students and use all subscales in the shortest version (5 items per scale except Worry with 7 items) for students in 12th grade.

This section of the report summarizes the analyses performed on the metacognitive instrument. We will report the results in three different sections as follows:

Part 1: the initial analyses on the 100-item instrument;
Part 2: analyses on the 70-item version of the instrument;
Part 3: analyses on the 50-item version of the instrument.

Part 1: 100-Item Instrument

The original instrument consisted of four subscales of metacognition (i.e., awareness, cognitive strategy, planning and self-checking; O'Neil, Baker, Jacoby,

Ni, & Wittrock, 1990); Effort; Curiosity; and Worry/Emotionality; it was administered to a group of 236 junior college students along with a 20-item math test (Kosmicki, 1993). There were two forms of the instrument: trait and state. The results of analyses will be presented first for the trait and then for the state form. The answers to all of the items in both forms, which were Likert-type items, ranged from 1 (*Almost Never*) to 4 (*Almost Always*) for the trait form and from 1 (*Not at All*) to 4 (*Very Much So*) for the state form. Traits were measured on a frequency dimension, whereas states were measured on an intensity dimension.

Results of Analyses for the 100-Item Trait Instrument

Table 1 presents the number of items, mean, standard deviation and Cronbach's Alpha coefficients for the subscales of 100-item trait instrument. As Table 1 indicates, the means ranged from 2.16 for Emotionality to 3.32 for Effort and standard deviations ranged from .43 for Effort to .80 for Emotionality. The reliability coefficients were relatively high for all of the subscales, ranging from .75 for Self Checking to .94 for Worry. The high reliability of some of the subscales was mainly due to the larger number of items and consistency between items. As seen in Table 1, for example, Worry with 23 items had an Alpha of .94, but Self-checking with only 7 items had an Alpha of .75.

Table 1

Number of Items, Mean, Standard Deviation and Cronbach's Alpha for the 100-item Trait Instrument

Variable	# of Items	Mean	SD	Alpha
AWARE	8	3.08	.53	.79
COGSTR	14	2.91	.49	.84
CURIOS	10	2.85	.63	.88
EFFORT	16	3.32	.43	.84
PLAN	9	3.06	.53	.83
SELFCHK	7	3.03	.53	.75
EMOTION	9	2.16	.80	.93
WORRY	23	2.29	.65	.94

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; CURIOS = Curiosity; EFFORT = Effort; PLAN = Planning; SELFCHK = Self-checking; EMOTION = Emotionality; WORRY = Worry.

To see individual item performance and to identify problematic items, that is, "attention" or "poor" items, several types of analyses were done on item level. Within each subscale, mean and standard deviation for each item were obtained. Also, correlation of each item with the total subscale score was computed to indicate the degree of fit of the particular item within the subscale. To get a comprehensive picture of how well the items fell within a subscale, a principal components analysis with varimax rotation was performed on the items within each subscale. This was done also to see if more than one category of item or factor existed under each subscale. Tables 2 through 9 present means, standard deviations and item-total correlations, as well as summary of the results of the principal components analyses (including factor loadings and communality for each item), for Awareness, Cognitive Strategy, Curiosity, Planning, Self-checking, Emotionality, Worry, and Effort respectively. As these tables indicate, individual items within and across subscales differ with respect to mean, standard deviation, item-total correlation, and factor loadings. In some subscales, such as Awareness, all items loaded on only one factor, whereas in some others, such as Cognitive Strategy, items loaded on more than one factor.

Table 2 summarizes the results of analyses for the Awareness subscale. As this table indicates, all items loaded on the first factor and all items were moderately correlated with the total Awareness score. The item-total correlation ranged from .41 for item 17 to .56 for items 29 and 35. The Alpha coefficient for this subscale was .79, which is acceptable but not high when compared with other subscales. The size of the item-total correlation and factor loadings for some of the items indicated that dropping those items might not have a large negative impact on the reliability of the scale and in some cases even would improve the reliability. For example, item 17 had lowest item-total correlation (.41) and lowest factor loading (.55). This item was placed under the "attention item" category and was dropped from the Awareness subscale without damaging the reliability of the subscale.

Similarly, Table 3 summarizes the results of analysis for the Cognitive Strategy subscale. There were 14 items in this subscale. The item means ranged from 2.22 for item 21 to 3.39 for item 12, and the standard deviation ranged from .73 for item 12 to .97 for item 21. Item-total correlation ranged from .20 for item 49 to .61 for item 22 and the Alpha coefficient for this subscale was .84. Unlike the Awareness subscale, items in this subscale loaded on more than

Table 2

SUBSCALE: Awareness (Trait) (N=236). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item Trait Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
4	0.69				3.07	.81	.55	.48
10	0.59				3.42	.78	.45	.35
17	0.55				2.93	.87	.41	.30
23	0.59				2.72	.86	.46	.35
29	0.69				3.02	.81	.56	.48
35	0.70				3.17	.81	.56	.49
40	0.68				3.28	.86	.55	.46
45	0.62				3.03	.83	.48	.39
EIG	3.31							
PC	41.30					Alpha = .79		

Note. R(IT) = Total item correlation; EIG = Eigenvalue; PC = Percent of variance.

one factor (three factors); however the eigenvalues and percent of variance extracted by each factor indicated that most of the items had relatively high loadings on the first factor. The percent of variance extracted by the first factor was 34.4 as compared with 10.0 and 8.1 for the second and third factors respectively. The fact the items within this subscale loaded on more than one factor and the low item-total correlation of some of the items in this subscale suggested that some items could be removed without having any negative impact on the reliability of the subscale. In fact, removing some the items might even increase the reliability. For example, item 49 had an item-total correlation of .20, no substantial loading on the first factor, and large loading on the third factor. All of these characteristics suggested putting this item in the "attention item" category. Similarly, item 15, with an item-total correlation of .37 and non-significant factor loading on the first factor, was removed. The same decision was made for item 50.

Table 3

SUBSCALE: Cognitive Strategy (Trait) (N=236). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item Trait Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
3	.70				3.31	.80	.51	.50
8	.69				3.11	.81	.56	.51
12	.66				3.39	.73	.47	.50
15		.48			2.80	.89	.37	.28
21		.77			2.22	.97	.45	.61
22	.48	.58			2.74	.90	.61	.56
28	.53	.53			2.66	.86	.60	.56
34	.55	.41			2.77	.81	.58	.48
39	.60				2.96	.87	.53	.44
44	.63				2.90	.85	.55	.47
48	.63				3.01	.86	.56	.49
49			.84		3.07	.93	.20	.71
50			.76		3.29	.80	.35	.63
52		.75			2.57	.95	.40	.58
EIG	4.82	1.40	1.14					
PC	34.4	10.0	8.10			Alpha = .84		

Similar results were obtained for the Curiosity subscale. These results are summarized in Table 4. This subscale had 10 items with an Alpha coefficient of .88. Means for these items ranged from 2.54 for item 96 to 3.30 for item 97. Standard deviations ranged from .82 for item 97 to 1.02 for item 96. Item-total correlations ranged from .34 for item 99 to .69 for item 93. Principal components analysis resulted in two factors for this subscale; however, the percent of variance extracted by the first factor was much higher than the second factor, that is, most of the items loaded highly on the first factor. Factor 1 extracted 47.9% and Factor 2 extracted 11.8% of the variance. The results of the analyses performed on items in this subscale, especially item-total correlations and factor

Table 4

SUBSCALE: Curiosity (Trait) (N=236). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item Trait Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
91	.76				2.79	.91	.55	.58
92	.83				2.62	.95	.63	.70
93	.66	.41			2.91	.87	.69	.60
94	.65				2.56	.92	.62	.53
95	.64				2.71	.91	.65	.56
96	.68				2.54	1.02	.61	.53
97		.81			3.30	.82	.53	.67
98		.84			3.18	.84	.59	.74
99		.69			3.00	.90	.34	.56
100	.43	.54			2.86	1.00	.61	.48
EIG	4.79	1.18						
PC	47.9	11.8				Alpha = .88		

loadings, were used to identify attention items. For example, item 97 with an item-total correlation of .53 and no significant loading on the first factor was marked as an "attention" item and was removed.

Table 5 summarizes the results of analyses for the trait Planning subscale. This subscale had 9 items with an Alpha coefficient of .83. The item means ranged from 2.06 for item 9 to 3.47 for item 1. Standard deviations ranged from .73 for item 1 to .92 for item 43. Item-total correlations for this subscale ranged from .32 for item 43 to .66 for item 38. For this subscale also, more than one factor was obtained (there were two factors with eigenvalues greater than 1 for this subscale). Like all the subscales discussed earlier with more than one factor, most of the items loaded highly on the first factor. The percent of variance extracted by the first factor was 43.6 as compared with 11.4% of variance extracted by the second factor. Summary statistics presented in Table 5 helped to identify and remove poor items. Item 43, for example, with low item-total correlation and no significant loading on the first factor, was

Table 5

SUBSCALE: Planning (Trait) (N=236). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item Trait Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
1	.51				3.47	.73	.48	.37
6		.47			3.14	.82	.47	.36
9	.76				2.06	.86	.58	.60
13	.66				3.22	.79	.52	.47
20		.73			2.83	.81	.64	.67
26	.42	.66			2.90	.86	.62	.61
32	.76				3.06	.83	.57	.60
38	.76				3.26	.74	.66	.65
43		.78			2.61	.92	.32	.61
EIG	3.90	1.00						
PC	43.6	11.4				Alpha = .83		

removed without having any negative effects on the reliability of the total scale. Similarly item 6 was labeled as an "attention" item and was removed.

The results of analyses for the Self-checking subscale are summarized in Table 6. This subscale had 7 items, and the Alpha coefficient for this subscale was .75. As Table 6 indicates, the item means ranged from 2.72 for item 16 to 3.41 for item 51. Standard deviations ranged from .76 for item 51 to .90 for item 16. All the items were moderately correlated with the total scale score. These correlations ranged from .41 for items 7 and 16 to .56 for item 27. Items were categorized under two factors, with the first factor extracting more variance than the second factor. The percent of variance extracted by the first factor was 40.3 and for the second factor was 14.5. The results of analyses suggested that items 7 and 16 could be marked for deletion because of relatively lower item-total correlation and low factor loading on the first factor.

Table 6

SUBSCALE: Self-checking (Trait) (N=236). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item Trait Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
2	.76				3.16	.86	.49	.61
7		.76			2.93	.84	.41	.59
14	.42	.57			3.04	.88	.53	.51
16		.73			2.72	.90	.41	.55
27	.63				3.05	.83	.56	.56
33		.48			2.93	.83	.42	.34
51	.82				3.41	.76	.43	.67
EIG	2.82	1.01						
PC	40.3	14.5				Alpha = .75		

Table 7 presents the results for the trait Emotionality subscale. There were 9 items in this subscale. The item means ranged from 1.89 for item 74 to 2.29 for items 65 and 73. Most of the items were highly correlated with the total scale score. The item-total correlations ranged from .65 to .81 and, as a result, the Alpha coefficient for this subscale was very high (.93). As one would expect, all items loaded highly on the first factor, and only one factor resulted. If there is a need to reduce the number of items for this subscale, one could easily remove items with lower item-total correlation, such as items 56 and 77.

The trait Worry subscale with 23 items is one of the most reliable subscales in the battery. The Alpha coefficient for the subscale was .94. Table 8 summarizes the results of analyses for this subscale. As Table 8 indicates, the item means ranged from 1.81 for item 62 to 3.37 for item 82, and item standard deviations ranged from .85 for item 82 to 1.08 for item 84. Most of the items were moderately to highly correlated with the total subscale score. Item-total correlations ranged from .25 for item 82 to .80 for item 76. The items in the Worry subscale loaded on three factors. The percent of variance extracted for the

Table 7

SUBSCALE: Emotionality (Trait) (N=236). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item Trait Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
56	.73				2.28	1.00	.66	.54
64	.79				2.28	.99	.73	.63
65	.86				2.29	1.07	.81	.73
67	.75				2.24	1.01	.68	.56
68	.84				2.28	1.01	.78	.70
72	.85				2.03	1.02	.80	.72
73	.83				2.29	1.05	.77	.69
74	.78				1.89	.98	.72	.61
77	.72				1.90	.96	.65	.51
EIG	5.70							
PC	63.3					Alpha = .93		

three factors were 45.4, 9.4, and 5.8 respectively. These figures indicated that most of the items loaded on the first factor. Based on the results summarized in Table 8, some of the items were removed from the Worry subscale without any major impact on the reliability of this subscale. For example, item 82, with a very low item-total correlation (.25) and non-significant loading on the first factor, was removed. With the same line of reasoning, item 81 was removed. Since items 81 and 82 have the highest loading on the third factor and only one other item (85) loaded moderately on this factor, removal of items 81 and 82 eliminated the third factor for this subscale. Removal of items 85 and 90 also did not have serious impact on the reliability of this scale.

The Effort subscale consisted of 16 items. Analyses done on this subscale are summarized in Table 9. As Table 9 indicates, the item-total correlations vary greatly from one item to other. One item (item 42) had a correlation of -.22 with the total and another item (item 31) had a correlation of .67 with the total scale

Table 8

SUBSCALE: Worry (Trait) (N=236). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item Trait Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
57	.65				2.67	1.04	.56	.51
58	.71				2.29	1.00	.59	.55
59	.77				2.05	.98	.66	.67
60	.69				1.96	1.06	.64	.56
61	.46	.52			2.34	.99	.66	.49
62	.76				1.81	.94	.64	.66
63	.75				2.08	.95	.69	.65
66	.78				1.83	.93	.69	.70
71	.65	.41			1.84	.86	.67	.59
75	.64	.50			2.01	1.05	.77	.66
76	.52	.70			2.00	1.03	.80	.76
79		.64			1.91	1.01	.59	.50
80	.63				2.19	.96	.74	.63
81			.77		3.09	.95	.39	.65
82			.82		3.37	.85	.25	.69
83		.61			2.26	1.07	.62	.50
84		.71			2.48	1.08	.75	.68
85		.60	.42		2.96	.93	.44	.54
86		.66			2.02	.99	.62	.59
87		.74			2.27	.99	.69	.65
88		.71			2.04	1.01	.75	.67
89		.59			2.47	.96	.66	.55
90		.57			2.91	.89	.49	.48
EIG	10.44	2.16	1.33					
PC	45.4	9.4	5.8			Alpha = .94		

Table 9

SUBSCALE: Effort (Trait) (N=236). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item Trait Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
5	.55	.44			3.78	.50	.43	.54
11	.74				3.52	.72	.58	.66
18			.71		3.37	.72	.31	.55
19		.82			2.94	1.00	.49	.69
24	.42		.52		3.20	.89	.61	.58
25	.43				3.44	.77	.59	.48
30	.71				3.55	.66	.64	.66
31		.62			3.27	.89	.67	.63
36	.71				3.50	.73	.63	.68
37	.70				3.81	.47	.36	.50
41			.57		3.22	.83	.54	.51
42				.72	2.19	.98	-.22	.60
46		.62			3.11	.82	.54	.51
47		.54			3.47	.75	.54	.45
53			.57		3.22	.88	.55	.55
54		.47	.58		3.58	.76	.37	.58
EIG	5.8	1.3	1.1	1.0				
PC	36.1	8.0	7.2	6.3		Alpha = .84		

score. The item means ranged from 2.19 for item 42 to 3.81 for item 37. The item standard deviations range from .47 for item 37 to 1.00 for item 19. Alpha reliability for this subscale was .84. The results of principal components analysis summarized in Table 9 indicated that the items in this subscale loaded on 4 factors; however, Factor 1 had most of the higher loadings. The percent of variance for Factor 1 was 36.1 as compared with 8.0, 7.2, and 6.3 for the second, third, and fourth factors respectively. Removal of item 42 with negative item-total correlation helped to improve reliability of this subscale. Item 18, with relatively low item-total correlation and no significant loading on the first or second factor, was also removed.

Results of Analyses for the 100-Item State Instrument

The state instrument was administered after students completed a math test (Kosmicki, 1993). The state instrument was similar to the trait instrument in number and format of items; also they both had same subscales. Table 10 presents the number of items, mean, standard deviation and Cronbach's Alpha coefficients for the subscales of the 100-item state instrument. Table 10 is comparable with Table 1, which reports similar data for the trait instrument. As Table 10 indicates, the subscale means range from 1.86 for the Worry subscale to 2.94 for Awareness. The subscale standard deviations range from .41 for Effort to .69 for Worry. The reliability levels for all the state subscales were acceptable and ranged from .77 for Self-checking to .90 for Worry.

Analyses were done on individual items under each category to see how items performed. Within each subscale, mean and standard deviation for each item were computed and correlation of each item with the total subscale score (i.e., item-total correlation) was obtained. The item-total correlation identified how well an item fit within a particular subscale. A principal components analysis with varimax rotation was also performed on the items within each subscale to see if items within any of the subscales were mutidimensional. The results of the item-level analyses are summarized in Tables 11 through 17 for Awareness, Cognitive Strategy, Curiosity, Planning, Self-checking, Worry, and Effort respectively. These results will be discussed for each of the subscales separately.

Table 11 shows means, standard deviations, item-total correlations, factor loadings, communalities, and reliability coefficient for the state Awareness subscale. As Table 11 indicates, the Awareness subscale had 8 items. The item means ranged from 2.70 for item 63 to 3.15 for item 48. The item standard deviations ranged from .87 for item 40 to 1.01 for item 21. Item-total correlations ranged from .33 for item 4 to .58 for items 28, 40 and 48. Alpha coefficient for this subscale is .78. Based on the summary results of the analyses done on items in the Awareness subscale, items 4 and 9 were omitted because they had relatively low item-total correlation (.33 and .34 respectively), and both of them had moderate loadings on the second factor. Thus, on the next version of the instrument, the Awareness subscale had only 6 items.

Table 10

Number of Items, Mean, Standard Deviation and Cronbach's Alpha for the 100-item State Instrument (N=210)

Variable	# of Items	Mean	SD	Alpha
AWARE	8	2.94	.58	.78
COGSTR	14	2.76	.53	.81
CURIOS	10	2.26	.68	.84
EFFORT	31	2.69	.41	.84
PLAN	9	2.90	.58	.80
SELFCHK	8	2.77	.63	.77
WORRY	14	1.86	.69	.90

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; CURIOS = Curiosity; EFFORT = Effort; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry.

Table 11

SUBSCALE: Awareness (State) (N = 210). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item State Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
4		.46			2.76	.93	.33	.25
9		.79			2.91	.97	.34	.63
21	.72				3.04	1.01	.39	.52
28	.75				3.13	.91	.58	.61
40	.44	.57			2.90	.87	.58	.52
48	.75				3.15	.96	.58	.62
53	.59				2.97	.95	.55	.50
63		.71			2.70	.94	.51	.57
EIG	3.15	1.07						
PC	39.4	13.4				Alpha = .78		

Similarly, Table 12 summarizes the results of analysis for the Cognitive Strategy subscale. In this category there were 14 items. The item means ranged from 2.00 for item 51 to 3.30 for item 7, and the standard deviations range from .87 for item 7 to 1.09 for item 60. Alpha reliability for this subscale was .81. The items in this subscale loaded on four factors, indicating that all the items within this subscale do not belong to the same category. By looking at the percent of variance extracted by each factor, however, it can be seen that most of the items had high loadings on the first factor. The percent of variance extracted by the first factor is 31.6 as compared with 10.0%, 8.1%, and 7.3% for the second, third and fourth factors respectively. Based on the results of analyses done on items within this category, the following items were removed: item 2, because of low item-total correlation (.39), low factor loading, and low communality; items 37, and 60, because of loading on the third factor. These two items mainly created Factor 3 for this subscale. Removal of these two items eliminated Factor 3 and created a more homogeneous set of items under the subscale. Item 51 was removed because of its negative item-total correlation. This item may belong to the Worry subscale. Item 26 was removed because it was very similar to item 55, and item 55 was kept.

Table 12

SUBSCALE: Cognitive Strategy (State) (N=210). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item State Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
2	.49				2.86	.89	.39	.33
7	.67				3.30	.87	.48	.57
13	.70				3.22	.88	.45	.52
26		.69			2.66	.93	.57	.59
34	.67				3.04	.95	.53	.52
37			.79		2.31	1.04	.40	.70
43	.66				2.83	1.07	.41	.47
47	.63	.41			3.10	.98	.63	.59
51				.90	2.00	1.08	-.008	.82
55		.79			2.80	.96	.53	.65
60			.85		2.51	1.09	.39	.77
66		.71			2.84	.97	.58	.60
67				.46	2.59	.99	.45	.47
75		.55			2.56	1.08	.35	.36
EIG	4.4	1.4	1.1	1.0				
PC	31.6	10.0	8.1	7.3		Alpha = .81		

Table 13 summarizes the results of analyses for the state Curiosity subscale. This subscale had 10 items and had an Alpha coefficient of .84. The item means ranged from 1.85 for item 100 to 3.12 for item 78. Standard deviations ranged from .91 for item 72 to 1.17 for item 76, and the item-total correlations ranged from .41 for item 76 to .67 for item 94. Items in this subscale loaded on two factors. Factor 1 explained 42.1% of the variance and Factor 2 explained 12.8% of the variance. Based on the analyses performed on items under this subscale, the following items were marked for deletion: item 76, because of relatively low item-total correlation (.41), and low factor loading on the first factor (.47); item 91, because it was very similar to item 94 and had

Table 13

SUBSCALE: Curiosity (State) (N=210). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item State Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
72	.61				2.08	.91	.49	.41
76	.47				2.19	1.17	.41	.28
77	.80				2.53	1.06	.63	.66
78	.67				3.12	.94	.45	.44
84	.75				2.73	1.03	.54	.57
88		.86			2.00	1.11	.53	.76
91	.45	.62			2.11	1.06	.64	.54
94	.62	.45			2.16	1.03	.67	.59
96	.55				2.17	.99	.53	.42
100		.88			1.85	1.09	.50	.77
EIG	4.2	1.3						
PC	42.1	12.8				Alpha = .84		

higher loading on the second factor; and item 100, because it was very similar in content to item 88.

The results of analyses for the Planning subscale with 9 items are presented in Table 14. As Table 14 indicates, the item means ranged from 2.13 for item 61 to 3.22 for item 39, and item standard deviations ranged from .87 for items 41 and 58 to 1.02 for item 61. Item-total correlations for this subscale ranged from .17 for item 61 to .62 for item 49. Items of this subscale loaded on two factors, Factor 1 explaining 41.3% of the variance and Factor 2 14.4% of the variance. The Alpha coefficient for this subscale was .80. The results of the analyses performed on the items and summarized in Table 14 suggested the omission of the following: item 5, because of relatively low item-total correlation (.38) and non-significant loading on the first factor; item 61, because of low item-total correlation (.17); and item 64, because of higher loading on the second factor.

Table 14

SUBSCALE: Planning (State) (N=210). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item State Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
5		.46			3.09	.96	.38	.28
14		.68			2.74	.91	.57	.59
23	.49	.48			2.69	.93	.54	.47
39	.80				3.22	.90	.59	.66
41	.86				3.16	.87	.58	.74
49	.74				3.09	.90	.62	.61
58	.74				3.24	.87	.59	.59
61		.75			2.13	1.02	.17	.60
64		.64			2.77	.99	.45	.47
EIG	3.72	1.30						
PC	41.3	14.4				Alpha = .80		

Table 15 summarizes the results of analyses for the state Self-checking subscale. As Table 15 shows, the Alpha coefficient for this subscale with 8 items is .77. The item means ranged from 2.65 for item 25 to 2.89 for item 35. Item standard deviations ranged from .92 for item 70 to 1.08 for item 35. Item-total correlations ranged from .38 for item 1 to .64 for item 31. Six items of this subscale loaded on the first factor and only two (item 19 and 57) loaded on the second factor. These two items, which also had relatively lower item-total correlation, were removed in order to increase internal consistency of the items.

Table 15

SUBSCALE: Self-checking (State) (N=210). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item State Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
1	.72				2.78	1.00	.38	.53
19		.78			2.85	.95	.39	.63
25	.54				2.65	1.05	.48	.38
31	.72				2.80	1.04	.64	.62
35	.63				2.89	1.08	.49	.45
46	.76				2.75	1.01	.59	.61
57		.78			2.75	.95	.40	.63
70	.46				2.72	.92	.43	.33
EIG	3.16	1.02						
PC	39.5	12.8				Alpha = .77		

The results of analyses for the state Worry subscale with 14 items are shown in Table 16. As Table 16 indicates, these item means were generally lower than item means of other subscales reported earlier. The item means for this subscale ranged from 1.40 for item 99 to 2.22 for item 79. Item variances, on the other hand, are generally higher than for other subscales and ranged from .83 for item 99 to 1.13 for item 95. Item-total correlations were moderate to high and ranged from .20 for item 81 to .73 for item 87. Alpha reliability for this subscale was .90. Items in this subscale loaded on three factors. Factor 1 explained 45.8% of the variance of the correlation matrix, Factor 2, 9.5%, and Factor 3, 7.6%. These percentages indicated that this subscale is mainly unidimensional, and, by removing a few of the items that loaded highly on the second and third factors, the internal consistency of the items could be increased even more.

Table 16

SUBSCALE: Worry (State) (N=210). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item State Instrument

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
20	.45		.54		2.01	1.08	.58	.52
73			.72		2.10	1.10	.59	.66
74		.60			2.00	1.06	.54	.48
79			.79		2.22	1.12	.55	.69
80	.72				1.66	.98	.68	.67
81	.41		.56		1.98	1.08	.20	.61
82	.85				1.49	.90	.60	.76
86		.59	.43		2.01	1.07	.46	.53
87	.54	.51			1.57	.88	.73	.63
89	.70				1.59	.96	.70	.66
93		.77			1.96	1.09	.65	.70
95		.69			2.13	1.13	.54	.60
97		.66			2.21	1.11	.50	.55
99	.81				1.40	.83	.67	.74
EIG	6.41	1.33	1.06					
PC	45.8	9.5	7.6			Alpha = .90		

The Effort subscale was the largest subscale of the state instrument. This subscale had 31 items. Table 17 presents the summary results of the analyses performed on this subscale. As Table 17 indicates, item means for this subscale ranged from 1.40 for item 36 to 3.50 for item 27. Item standard deviations ranged from .78 for item 36 to 1.08 for items 3, 52, 54, and 12. Item-total correlations were very different across the items. For some items there were negative item-total correlations and for some others there were relatively high positive correlations. The range of item-total correlation for this subscale was from -.06 for item 3 to .73 for item 33. The Alpha coefficient for this subscale was .84. The items of this subscale loaded on 7 factors. The percents of variance explained by these 7 factors were 29.4%, 9.7%, 5.7%, 4.5%, 3.9%, and 3.6%

respectively. The results of the analyses performed on the items of this subscale suggested that several items could be omitted without having any negative impact on the reliability of the scale. The removal of some items which had negative correlation with the total scale score even improved the reliability of the scale. Based on the results summarized in Table 17, the following items were omitted: item 3, because of negative (near zero) item-total correlation (-.06) and loading on the sixth factor; item 8, because of relatively low loading and loadings on different factors; item 11, because of low item-total correlation and loading on the fourth factor (this item did not seem to belong to Effort subscale); item 17, because of low item-total correlation (.09); item 36, because of low item-total correlation (.09); item 44, because of low item-total correlation (.09); item 56, because of low item-total correlation (.10) (this item did not seem to belong to Effort subscale); item 62, because of negative item-total correlation (-.29); item 68, because of negative (near zero) item-total correlation (-.05); and item 50, because it seemed to fit more in the Cognitive Strategy category.

Table 17

SUBSCALE: Effort (State) (N = 210). Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the 100-item State Instrument

	Factor Loadings										
Item#	F1	F2	F3	F4	F5	F6	F7	Mean	SD	R(IT)	COMM
3						.62		2.12	1.08	-.06	.50
6							.58	1.82	1.01	-.21	.63
8		.41		.53				2.42	.97	.30	.60
10	.74							3.40	.91	.51	.66
11				.74				2.46	1.03	.29	.63
12								2.98	1.08	.52	.50
15		.56						3.15	.91	.46	.52
16		.57						2.76	.93	.45	.62
17			.56					1.71	.93	.09	.40
18	.72							2.96	1.00	.49	.52
22							.82	2.92	1.06	.17	.71
24	.57							3.14	.98	.58	.51
27	.76							3.50	.80	.65	.70
29		.66						2.34	1.07	.22	.57
30	.77							3.25	.93	.70	.72
32	.77							3.17	.97	.66	.76
33	.65							3.25	.94	.73	.68
36			.75					1.40	.78	.09	.67
38					.78			3.15	.93	.30	.69
42	.40	.41			.40			3.09	1.03	.54	.50
44			.59	.48				1.54	.88	.09	.67
45	.74							3.20	.93	.58	.63
50	.57							3.28	.97	.49	.52
52	.48	.47						3.00	1.08	.56	.54
54		.63						3.00	1.08	.54	.60
56			.68					1.55	.90	.10	.58
59		.61						2.92	.99	.48	.57
62						.65		2.00	1.11	-.29	.58
65		.68						2.96	1.02	.56	.67
68			.65					1.80	1.02	-.05	.56
69	.50	.45						3.11	1.03	.54	.54
EIG	9.10	39.0	1.78	1.39	1.2	1.1					
PC	29.4	9.7	5.7	4.5	3.9	3.6				Alpha = .84	

Part 2: 70-Item Instrument

Results of Analyses for the 70-Item State Instrument

Table 18 presents the number of items, mean, standard deviation and Cronbach's Alpha coefficients for the subscales of 100-item state instrument. After removing 30 poor items from different subscales, the same type of analyses done on the 100-item instrument were repeated for the reduced form of 70 items. This analysis was done on the same data set (Kosmicki, 1993) discussed in the prior section. Item means, item standard deviations and item-total correlations were computed for items for each subscale. Alpha coefficients were also obtained for each of the subscales. Also, principal components analysis was performed on items within each subscale to see how removing poor items affected the dimensionality of the subscales. For each of the subscales in the reduced instrument, the same summary tables were generated. Tables 19 through 25 present the results of analyses on the state instrument subscales after removing poor items. The type of data presented in these tables and the format of the tables are identical with the Tables 11 to 17 to facilitate cross comparisons of the data before and after removing poor items. For example, Table 11 which summarizes the results of analyses for the Awareness subscale is comparable with Table 19 which presents the same results for the Awareness subscale after removing poor items. We will not present as much detail nor describe the results of analyses on the short form as extensively we did for the original form. We ask those readers who are interested in the detailed analyses to compare the two sets of tables. We rather prefer to compare the subscales with respect to their number of items, number of factors, and reliabilities before and after removing poor items. Table 26 provides such information.

Table 18

Number of Items, Mean, Standard Deviation and Cronbach's Alpha for State Short Version

Variable	# of Items	Mean	SD	Alpha
AWARE	6	2.98	.61	.79
COGSTR	8	3.00	.55	.81
CURIOS	7	2.40	.65	.81
PLAN	5	3.08	.51	.83
SELFCHK	5	2.77	.50	.75
WORRY	11	1.82	.68	.90
EFFORT	17	3.04	.54	.90

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; CURIOS = Curiosity; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort.

Table 19

SUBSCALE: Awareness (State) Short Version. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the State Short Version

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
21	.56				3.04	1.01	.40	.32
28	.74				3.13	.91	.61	.55
40	.70				2.90	.87	.56	.50
48	.76				3.15	.96	.62	.58
53	.72				2.97	.95	.57	.52
63	.63				2.70	.94	.48	
EIG	2.87							
PC	47.8					Alpha = .79		

Table 20

SUBSCALE: Cognitive Strategy (State) Short Version. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the State Short Version

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
2	.54				2.86	.89	.41	.30
7	.71				3.30	.87	.57	.51
13	.65				3.22	.88	.51	.42
34	.67				3.04	.95	.54	.44
43	.57				2.83	1.07	.44	.31
47	.76				3.10	.98	.64	.58
55	.58				2.80	.96	.46	.34
66	.66				2.84	.97	.55	.44
EIG	3.35							
PC	41.8					Alpha = .81		

Table 21

SUBSCALE: Curiosity (State) Short Version. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the State Short Version

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
72	.68				2.08	.91	.53	.68
77	.75				2.53	1.06	.61	.75
78	.62				3.12	.94	.48	.62
84	.72				2.73	1.03	.58	.72
88	.54				2.00	1.11	.41	.54
94	.76				2.16	1.03	.64	.76
96	.68				2.17	.99	.54	.68
EIG	3.28							
PC	46.8					Alpha = .81		

Table 22

SUBSCALE: Planning (State) Short Version. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the State Short Version

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
23	.62				2.69	.93	.46	.38
39	.81				3.22	.90	.69	.66
41	.84				3.16	.87	.72	.70
49	.79				3.09	.90	.66	.62
58	.78				3.24	.87	.63	.60
EIG	2.97							
PC	59.4					Alpha = .83		

Table 23

SUBSCALE: Self Checking (State) Short Version. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the State Short Version

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
1	.65				2.78	1.00	.44	.43
25	.66				2.65	1.05	.46	.43
31	.79				2.80	1.04	.63	.62
35	.67				2.89	1.08	.45	.44
46	.77				2.75	1.01	.58	.59
EIG	2.5							
PC	50.4					Alpha = .75		

. Table 24

SUBSCALE: Worry (State) Short Version. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the State Short Version

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
20	.66				2.01	1.08	.58	.49
74	.61	.41			2.00	1.06	.54	.54
79	.56				2.22	1.12	.49	.32
80	.77				1.66	.98	.70	.68
81	.75				1.98	1.08	.69	.56
82	.73				1.49	.90	.64	.68
87	.79				1.57	.88	.72	.63
89	.79				1.59	.96	.72	.64
93	.69	.52			1.96	1.09	.63	.75
95	.62	.50			2.13	1.13	.54	.63
99	.78				1.40	.83	.69	.69
EIG	5.5	1.0						
PC	50.4	9.7				Alpha = .90		

Table 25

SUBSCALE: Effort (State) Short Version. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the State Short Version

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
6	-.51				1.82	1.01	-.38	.41
15		.54			3.15	.91	.57	.47
24	.51				3.14	.98	.64	.51
27	.73				3.50	.80	.75	.71
32	.78				3.17	.97	.77	.76
42			.62		3.09	1.03	.57	.52
52			.45		3.00	1.08	.57	.43
54			.59		3.00	1.08	.57	.53
59		.83			2.92	.99	.55	.75
65		.78			2.96	1.02	.61	.71
10	.73				3.40	.91	.63	.64
16			.82		2.76	.93	.43	.68
18	.64				2.96	1.00	.54	.48
30	.72				3.25	.93	.77	.72
33	.54		.45		3.25	.94	.75	.63
45	.71				3.20	.93	.68	.61
69	.47	.60			3.11	1.03	.65	.60
EIG	7.8	1.2	1.1					
PC	46.1	7.3	6.4			Alpha = .90		

As Table 26 shows, removing poor items in most cases increased the reliability of the subscale and reduced the number of items to a more manageable level. There were originally 94 items (100 items minus 6 Emotionality items) in the instrument. From the total items, 35 items (about 37% of the original items) were removed, yet the average reliabilities increased from .82 to .83. The difference between .82 and .83 may not be substantial, but at least it suggests that the reduction of items by 37% did not have any negative impact on the reliability of the instrument. By looking at the reliability of

Table 26

Number of Items, Number of Factors, and Alpha Coefficients for the Full and the Reduced State Instrument

	Number of Items		Number of Factors		Alpha	
Subscale	Full	Reduced	Full	Reduced	Full	Reduced
AWARE	8	6	2	1	.78	.79
COGSTR	14	8	4	1	.81	.81
CURIOS	10	7	2	1	.84	.81
PLAN	9	5	2	1	.80	.83
SELFCHK	8	5	2	1	.77	.75
WORRY	14	11	3	2	.90	.90
EFFORT	31	17	7	3	.84	.90
EMOTION[a]	N/A	N/A	N/A	N/A	N/A	N/A

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; CURIOS = Curiosity; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort; EMOTION = Emotionality.

[a] Emotionality subscale is not included.

individual subscales, it is apparent that the reliabilities of the long form and the short form are almost identical except in few cases. For the Curiosity subscale, the shorter form is a little less reliable than the longer form; however, the difference is not statistically significant (.84 for the long form and .81 for the short form, $z = .41$, $p > .05$) (Edwards, 1961, pp. 304-306). The Effort subscale, on the other hand, gained reliability after omitting poor items. The Alpha coefficient for the Effort subscale in the long form was .84 and in the short form, after losing almost half of its items, was .90. Another point in Table 26 regarding the efficiency of the short versus the long form is the reduction in number of factors in the short form. Principal components analyses yielded 2, 3, 4 and even 7 factors for many of the subscales of the long form. The minimum number of factors for the long form was 2. That is, the items under subscales in the long form were not unidimensional. The problem of multidimensionality of items in the long form created difficulties when computing subscale scores. In the short form however, this problem was reduced considerably. Items under five of the seven subscales loaded on only one factor in the short form as

compared with two or more factors in the long form. For example, items under Cognitive Strategy in the long form loaded on four factors which indicated that in this subscale, items were under four different categories. This clearly created problem in obtaining composite score for this subscale. Reducing the number of items from 14 to 8 did not have any effect on the reliability, but caused the items to be grouped under one category, that is, a more homogeneous set of items resulted. In the Effort subscale, as Table 25 indicates, after reducing the number of items by 55%, reliability was increased from .84 to .90 and number of factors decreased from 7 to 3.

In summary, after identifying and removing the poor items, the resulting instrument had more homogeneous items within the subscales and was easier to administer.

We decided to use the state short form on another group of subjects to examine the psychometric properties of the items and cross validate the previous findings. Due to time constraints for administration, there was a need to reduce the number of items even further. Thus, we looked again at results of analyses done on items under each subscale, and we identified some additional marginal items which could be removed without having a significant impact on the reliability of the instrument. On the second review of the items, 12 items were identified as "marginal" items and were removed, 5 new items were added to the Planning subscale, and 3 new items were added to the Self-checking subscale. Finally, the Curiosity subscale was eliminated. As a result of these changes, a 48-item instrument resulted. The items removed from Worry were 81, 89, and 99, and the items removed from Effort were 6, 32, 33, 42, 54, 59, 65, and 69. This state instrument was administered to another group of 230 high-school students (Khabiri, 1993). Means and standard deviations as well as Alpha coefficients for each of the subscales were computed and principal components analysis with varimax rotation was applied on the subscale items to see how items grouped together under each subscale. Table 27 reports number of items, mean, standard deviation, and Alpha coefficient for each of the six subscales. As Table 27 indicates the subscale means ranged from 1.74 for Worry to 2.81 for Effort, and subscale standard deviations ranged from .54 for Cognitive Strategy and Planning to .61 for Effort. Alpha coefficients ranged from .70 for Awareness and Worry to .82 for Effort. The Alpha coefficients for some of the subscales

Table 27

Number of Items, Mean, Standard Deviation and Cronbach's Alpha for High School Students Prior to Pilot Study Before Deletion

Variable	# of Items	Mean	SD	Alpha
AWARE	6	2.57	.58	.70
COGSTR	8	2.56	.54	.71
PLAN	10	2.28	.54	.81
SELFCHK	8	2.40	.60	.80
WORRY	7	1.74	.55	.70
EFFORT	9	2.81	.61	.82

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort.

were low. For example, the Alpha coefficients for the subscales Awareness, Cognitive Strategy, and Worry were around .70, a minimally acceptable level.

Tables 28 through 33 summarize the results of analyses done on the item level for each of the subscales. These tables are comparable with the previous tables summarizing the results of the longer version of the instrument. Readers who are interested in comparing the performance of individual items on different groups can compare these tables. Based on the results of analyses presented in Tables 28 through 33, poor items were identified and removed to determine how their removal would affect the reliability of the instrument. Out of the 48 items in the reduced form, 7 items (15%) were marked as "attention items" and were deleted. The following items were removed: item 21 from Awareness, item 2 from Cognitive Strategy, one of the newly-added items from Planning, item 1 from Self-checking, item 74 from Worry, and items 16 and 18 from Effort. Mean and standard deviation for each item under each of the subscales were computed. Principal components analysis was performed on the subscale items and Alpha coefficient was obtained for each subscale of the 41-item instrument. Table 34 summarizes the descriptive statistics for this form. As this table indicates, subscale means ranged from 1.69 for Worry to 2.88 for Effort, and

Table 28

SUBSCALE: Awareness for High School Students Prior to Pilot Study Before Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
21	.47				2.83	.98	.29	.22
28	.63				2.86	.84	.43	.39
40	.62				2.37	.90	.41	.38
48	.77				2.70	.93	.58	.59
53	.64				2.46	.92	.43	.41
63	.68				2.17	.92	.47	.46
EIG	2.44							
PC	40.7					Alpha = .70		

subscale standard deviations ranged from .56 for Planning to .67 for Effort. After removing 7 poor items from the reduced form, the reliability of the subscales (Alpha coefficients) stayed the same or even increased in some cases. The Alpha coefficients for this form (41-item form) ranged from .71 for Awareness and Cognitive Strategy to .83 for Effort.

Table 29

SUBSCALE: Cognitive Strategy for High School Students Prior to Pilot Study Before Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
2	.48				2.29	.89	.26	.23
7		.84			3.15	.91	.33	.70
13		.72			3.04	.95	.38	.54
34	.59				2.37	.97	.46	.43
43	.67				2.43	1.02	.47	.48
47	.67				2.46	.96	.43	.46
55	.45	.50			2.44	.89	.47	.45
66	.67				2.32	.90	.40	.45
EIG	2.66	1.10						
PC	33.2	13.7				Alpha = .71		

Table 30

SUBSCALE: Planning for High School Students Prior to Pilot Study Before Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
**a	.45				1.67	.89	.42	.31
**a		.62			2.79	.91	.52	.47
23	.69				1.91	.82	.56	.55
**a	.61				2.11	.92	.57	.51
39	.57				2.45	1.00	.53	.44
41	.71				2.10	.91	.47	.52
**a		.65			2.52	.85	.51	.49
49		.77			2.62	.83	.45	.59
58		.74			2.76	.94	.48	.56
**a	.61				1.86	.93	.27	.39
EIG	3.65	1.17						
PC	36.5	11.7				Alpha = .80		

[a] ** = Denotes new items. (The remaining items were from the 100-item State/Trait Instrument.)

Table 31

SUBSCALE: Self-checking for High School Students Prior to Pilot Study Before Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
1	.80				2.39	.88	.41	.66
**[a]	.56	.45			2.55	.95	.59	.51
**	.64	.41			2.26	.92	.64	.58
25	.54				2.19	.99	.48	.39
31		.74			2.32	1.01	.52	.62
35		.83			2.81	.92	.43	.70
46	.66				2.48	.97	.55	.50
**	.56				2.19	.89	.43	.35
EIG	3.29	1.03						
PC	41.1	12.8				Alpha = .80		

[a] **Denotes new items. (These items were added to the pool for the 100-item State/Trait Instrument.)

Table 32

SUBSCALE: Worry for High School Students Prior to Pilot Study Before Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
74		.86			2.03	.93	.21	.76
20	.44				1.50	.89	.43	.39
79	.57				1.95	1.09	.41	.37
80	.71				1.65	.91	.56	.58
82	.78				1.43	.79	.44	.61
87	.78				1.52	.81	.50	.61
93		.61			2.11	1.02	.38	.45
EIG	2.63	1.14						
PC	37.7	16.3				Alpha = .70		

Table 33

SUBSCALE: Effort for High School Students Prior to Pilot Study Before Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
18	.58				2.82	.88	.43	.34
10	.81				3.06	.91	.67	.66
15	.68				2.87	.88	.57	.49
16		.91			2.33	.97	.23	.83
24	.42	.60			2.33	1.08	.50	.54
27	.78				3.13	.94	.72	.69
30	.77				2.84	.95	.63	.62
45	.60				3.09	.92	.48	.37
52	.59				2.86	1.03	.50	.37
EIG	3.87	1.04						
PC	43.0	11.6				Alpha = .82		

Table 34

Number of Items, Mean, Standard Deviation and Cronbach's Alpha for High School Students Prior to Pilot Study After Deletion

Variable	# of Items	Mean	SD	Alpha
AWARE	5	2.51	.61	.71
COGSTR	7	2.60	.57	.71
PLAN	9	2.33	.56	.81
SELFCHK	7	2.40	.63	.79
WORRY	6	1.69	.60	.72
EFFORT	7	2.88	.67	.83

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort.

Tables 35 through 40 summarize the results of item-level analyses for the subscales of the 41-item instrument. As mentioned earlier, the structure of the tables reporting item-level analyses are similar to facilitate cross-form comparisons. For example, Tables 28-33 are comparable with Tables 35 through 40. The only difference is that in the latter tables, there are fewer items because the "attention items" have been removed. Readers who are interested in comparing item statistics before and after "attention items" were removed can compare the two sets of tables.

Table 35

SUBSCALE: Awareness for High School Students Prior to Pilot Study After Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
28	.62				2.86	.84	.41	.39
40	.64				2.37	.90	.42	.41
48	.77				2.70	.93	.56	.59
53	.66				2.46	.92	.45	.44
63	.70				2.17	.92	.48	.48
EIG	2.30							
PC	46.0					Alpha = .71		

Table 36

SUBSCALE: Cognitive Strategy for High School Students Prior to Pilot Study After Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
7		.84			3.15	.91	.33	.70
13		.74			3.04	.95	.37	.57
34	.66				2.37	.97	.48	.48
43	.69				2.43	1.02	.45	.49
47	.70				2.46	.96	.43	.49
55	.50	.45			2.44	.89	.48	.46
66	.66				2.32	.90	.38	.44
EIG	2.55	1.09						
PC	36.5	15.5				Alpha = .71		

Table 37

SUBSCALE: Planning for High School Students Prior to Pilot Study After Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
**[a]	.46				1.67	.89	.41	.30
**		.62			2.79	.91	.52	.48
23	.73				1.91	.82	.55	.57
**	.72				2.11	.92	.59	.58
39	.61				2.45	1.00	.52	.45
41	.76				2.10	.91	.45	.58
**		.61			2.52	.85	.52	.48
49		.78			2.62	.83	.46	.62
58		.75			2.76	.94	.50	.58
EIG	3.55	1.10						
PC	39.5	12.2				Alpha = .81		

[a] **Denotes new items. (These items were added to the item pool in the 100-item State/Trait Instrument.)

Table 38

SUBSCALE: Self Checking for High School Students Prior to Pilot Study After Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
**[a]	.71				2.55	.95	.56	.50
**	.76				2.26	.92	.62	.58
25	.63				2.19	.99	.48	.39
31	.69				2.32	1.01	.54	.47
35	.60				2.81	.92	.46	.35
46	.67				2.48	.97	.52	.46
**	.56				2.19	.89	.42	.31
EIG	3.06							
PC	43.7					Alpha = .79		

[a] **Denotes new items. (These items were not yet added to the item pool for the 100-item State/Trait Instrument.)

Table 39

SUBSCALE: Worry for High School Students Prior to Pilot Study After Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
20	.58				1.50	.89	.41	.34
79	.62				1.95	1.09	.43	.38
80	.76				1.65	.91	.57	.58
82	.70				1.43	.79	.49	.49
87	.74				1.52	.81	.56	.55
93	.50				2.11	1.02	.72	.25
EIG	2.58							
PC	43.0					Alpha = .72		

Table 40

SUBSCALE: Effort for High School Students Prior to Pilot Study After Deletion. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for High School Students

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
10	.80				3.06	.91	.68	.64
15	.70				2.87	.88	.57	.49
24	.60				2.33	1.08	.47	.36
27	.85				3.13	.94	.75	.72
30	.81				2.84	.95	.68	.65
45	.60				3.09	.92	.48	.35
52	.60				2.86	1.03	.48	.36
EIG	3.57							
PC	51.0					Alpha = .83		

Table 41 compares the full 100-item instrument with the 41-item version. Note that an additional 5 new Planning items and 3 new Self-checking items were added to the item pool. As Table 41 indicates, the number of items for most of the subscales was reduced substantially in the new form. Effort, with the highest number of items in the original form (31 items), lost most of its items and was reduced to a 7-item subscale; however, the reliability of this subscale in the original form with 31 item is almost identical with the reliability of this subscale with only 7 items (.84 in the full versus .83 in the reduced form). In some other subscales, however, the Alpha coefficient dropped considerably. In the Cognitive Strategy subscale, for example, the Alpha decreased from .81 to .71 when the number of items in the subscale was reduced.

A comparison of the 100-item original instrument with the reduced form may not be valid because the statistics were based on two different groups of subjects (junior college students vs. high school students), which may represent two different populations. Thus, any difference in the size of Alpha may be attributable to initial differences between the two groups. However, because very similar results were obtained on the subscales with about the same number of items in the full and the reduced forms, the two groups of subjects may be considered as being drawn from the same population.

Table 41

Number of Items, Number of Factors and Alpha Coefficients for the Full 100-Item State Instrument and the Reduced State Scale (41 items)

	Number of Items		Number of Factors		Alpha	
Subscale	Full	Reduced	Full	Reduced	Full	Reduced
AWARE	8	5	2	1	.78	.71
COGSTR	14	7	4	2	.81	.71
CURIOS[a]	10	N/A	2	N/A	.84	N/A
PLAN	9	9	2	2	.80	.81
SELFCHK	8	7	2	1	.77	.79
WORRY	14	6	3	1	.90	.72
EFFORT	31	7	7	1	.84	.83

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; CURIOS = Curiosity; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort.

[a] Curiosity was not included in Khabiri (1993).

The reduced form of the scale with six subscales was used on 376 8th-grade and 464 12th-grade students in the pilot phase of this experimental motivation study with two modifications. The number of items for the Self-checking subscale was increased from 7 to 11 and the Curiosity subscale was put back into the battery. The 50-item instrument was placed following the math tests, at the end of booklets prepared for 8th- and 12th-grade students. The booklets contained some NAEP background variables initially, NAEP Block 3 and 7 math items, and the 50-item metacognitive instrument. Since there was not enough time to complete the booklets and because the metacognitive questions were placed at the end of the booklets, there were many unanswered items, especially for the 8th-grade pilot students. Because of this problem, our analysis in this Appendix was performed only on the 12th-grade pilot data. Table 42 summarizes the descriptive statistics for the subscales used on the 12th-grade pilot study. As Table 42 indicates, the subscale means ranged from 1.63 for the Worry subscale to 2.87 for Effort. Subscale standard deviations ranged from .68 for Worry to .90 for Awareness. Subscale reliabilities ranged from .77 for Worry to .87 for Self-checking.

Table 42

Number of Items, Mean, Standard Deviation and Cronbach's Alpha for the Pilot Study, 12th Grade

Variable	# of Items	Mean	SD	Alpha
AWARE	5	2.54	.90	.82
COGSTR	7	2.58	.82	.83
PLAN	9	2.38	.75	.84
SELFCHK	11	2.37	.73	.87
WORRY	6	1.63	.68	.77
EFFORT	7	2.87	.80	.84
CURIOS	5	1.97	.79	.78

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort; CURIOS = Curiosity.

The analyses performed on items within the subscales are summarized in Tables 43 through 49. Again, these tables are comparable with those reporting the results of item-level analyses for the original (full) and reduced forms.

Table 43

SUBSCALE: Awareness for Pilot 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Pilot Study, 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
28	.66				3.12	.97	.51	.44
40	.68				2.54	1.08	.52	.46
48	.83				2.55	1.31	.70	.69
53	.86				2.50	1.28	.74	.74
63	.79				1.98	1.20	.64	.62
EIG	2.94							
PC	58.8					Alpha = .82		

Table 44

SUBSCALE: Cognitive Strategy for Pilot 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Pilot Study, 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
7		.75			3.02	.97	.45	.58
13		.79			2.97	1.01	.51	.64
34		.71			2.62	1.07	.62	.64
43	.41	.53			2.52	1.19	.52	.45
47	.77				2.40	1.24	.68	.70
55	.86				2.31	1.28	.64	.76
66	.85				2.20	1.31	.63	.76
EIG	3.48	1.04						
PC	49.8	14.9				Alpha = .83		

Table 45

SUBSCALE: Planning Pilot for 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Pilot Study, 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
**[a]	.69				1.99	1.05	.40	.48
**[a]	.58				2.93	.98	.57	.47
23	.77				1.97	.98	.56	.61
**[a]	.78				2.26	1.06	.53	.62
39	.67				2.56	1.12	.58	.53
41	.62				2.16	1.12	.51	.45
**[a]		.84			2.49	1.21	.62	.75
49		.90			2.52	1.27	.61	.83
58		.87			2.58	1.37	.56	.78
EIG	3.93	1.58						
PC	43.7	17.6				Alpha = .84		

[a] **Denotes new items that were initially introduced with the High School Students Prior to the Pilot Study and that were carried over to the Pilot Study, 12th Grade.

Table 46

SUBSCALE: Self-checking for Pilot 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Pilot Study, 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
@[a]	.76				2.20	1.04	.58	.60
@[a]	.70				2.14	1.07	.48	.49
@[a]	.41				2.24	1.08	.48	.32
@[a]		.89			2.50	1.39	.46	.80
**[b]	.69				2.61	.95	.58	.53
**[b]	.74				2.39	1.01	.62	.59
25	.65				2.52	1.11	.54	.47
@[a]	.68	.42			2.46	1.10	.71	.64
35		.63			2.74	1.12	.62	.55
46		.60			2.22	1.12	.57	.48
**[b]		.80			2.03	1.25	.61	.69
EIG	4.83	1.35						
PC	43.9	12.3				Alpha = .87		

[a] @ Denotes new items that were introduced for the Pilot Study.

[b] ** Denotes new items that were initially introduced with the High School Students Prior to the Pilot Study and that were carried over to the Pilot Study, 12th Grade.

Table 47

SUBSCALE: Worry for Pilot 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Pilot Study, 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
20	.67				1.56	.96	.51	.45
79	.68				2.23	1.24	.50	.46
80	.78				1.56	.97	.62	.61
82	.79				1.31	.79	.62	.62
**[a]	.77				1.39	.83	.61	.59
93	.47				1.75	1.17	.32	.22
EIG	2.96							
PC	49.4					Alpha = .77		

[a] ** Denotes new items that were initially introduced with the High School Students Prior to the Pilot Study and that were carried over to the Pilot Study, 12th Grade.

Table 48

SUBSCALE: Effort for Pilot 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Pilot Study, 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
10	.84				3.07	1.00	.63	.71
15	.82				2.92	.97	.68	.70
24	.67				2.58	1.13	.54	.48
27	.87				3.22	.97	.76	.81
30	.79				2.79	1.07	.71	.70
45		.82			2.92	1.19	.55	.75
52		.89			2.55	1.41	.42	.80
EIG	3.82	1.12						
PC	54.6	16.1				Alpha = .84		

Table 49

SUBSCALE: Curiosity for Pilot 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Pilot Study, 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
72	.75				1.99	.97	.58	.56
77	.77				2.20	1.05	.59	.59
88	.63				1.69	.97	.45	.39
94	.78				2.18	1.20	.63	.61
96	.73				1.77	1.17	.56	.54
EIG	2.69							
PC	53.8					Alpha = .78		

Main Study

Since many of the 8th-grade pilot study group and some of the 12th-grade pilot study group could not answer all the metacognitive questions, we decided to reduce the number of items even further based on the pilot study results and based on the NCES staff input on item sensitivity. We reduced the number of items in all of the subscales to 5, except for the Worry subscale which had 8 items. As indicated earlier, the percentage of unreached items for 8th-grade students was higher; therefore, we needed to develop a shorter version of the instrument for the 8th-grade group. Since having fewer than 5 items in each subscale affected the subscale reliability dramatically, we decided to use fewer subscales for the 8th-grade main study rather than having fewer than 5 items in each subscale. Therefore, two different versions of the instrument were prepared. For the 12th-grade students a six-subscale version was used. Five of the subscales in this version (Awareness, Cognitive Strategy, Planning, Self-checking and Effort) had 5 items each and one subscale (Worry) had 8 items. For the 8th-grade students, a version with four subscales was used. The subscales for the 8th-grade group were: Cognitive Strategy, Self-checking, and Effort each with 5 items and Worry with 8 items. Over 95% of both 8th- and

12th-grade students in the main study sample answered all the questions in the booklets. Table 50 summarizes the results of analyses of the four subscales for the 8th-grade students in the main study. As Table 50 indicates, the subscale means ranged from 1.75 for Worry to 3.38 for Effort, and the subscale standard deviations ranged from .62 for Worry to .65 for Cognitive Strategy. Alpha coefficients for 8th-grade students on two of the four of the subscales were low. The Alpha coefficient for Cognitive Strategy was .61, for Self-checking was .64, for Worry was .79, and for Effort was .76. The low reliability of the subscales for the 8th-grade students was mainly due to low variability of the responses. Tables 51 through 54 present the summary of the item-level analyses for 8th-grade students on Cognitive Strategy, Self-checking, Worry, and Effort respectively. These results are comparable with the results obtained on the original instrument and the reduced forms reported earlier.

Table 50

Number of Items, Mean, Standard Deviation and Cronbach's Alpha for the Main Study 8th Grade

Variable	# of Items	Mean	SD	Alpha
COGSTR	5	2.75	.65	.61
SELFCHK	5	2.68	.63	.64
WORRY	8	1.75	.62	.79
EFFORT	5	3.38	.63	.76

Note. COGSTR = Cognitive Strategy; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort.

Table 51

SUBSCALE: Cognitive Strategy for Main Study 8th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, and Cronbach's Alpha for the Main Study, 8th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
34	.60				2.61	1.04	.34	.36
43	.59				2.72	1.08	.33	.34
47	.61				2.89	1.07	.36	.38
55	.65				2.77	1.00	.39	.42
66	.69				2.78	.95	.43	.48
EIG	1.98							
PC	39.6					Alpha = .61		

Table 52

SUBSCALE: Self Checking for Main Study 8th Grade Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study, 8th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
31	.69				2.77	.98	.43	.48
35	.64				2.95	.98	.39	.41
**[a]	.50				2.35	1.04	.29	.25
46	.62				2.46	.98	.37	.38
**[a]	.76				2.86	.92	.51	.57
EIG	2.09							
PC	41.8					Alpha = .64		

[a] ** = Denotes new items that were initially introduced with the High School Students Prior to the Pilot Study and that were carried over to the Main Study, 8th Grade.

Table 53

SUBSCALE: Worry for Main Study 8th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study, 8th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
20	.54	.40			1.77	.96	.53	.46
82	.78				1.41	.82	.42	.62
89	.71				1.52	.92	.52	.54
80	.71				1.61	.96	.58	.59
87	.41	.47			1.55	.88	.48	.39
95		.83			1.89	1.04	.46	.69
79		.51			2.40	1.15	.54	.41
93		.77			1.88	1.05	.52	.63
EIG	3.29	1.03						
PC	41.1	12.9				Alpha = .79		

Table 54

SUBSCALE: Effort for Main Study 8th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study, 8th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
15	.69				3.24	.88	.50	.47
30	.79				3.34	.90	.61	.63
10	.82				3.39	.89	.65	.67
45	.52				3.35	.94	.35	.37
27	.77				3.59	.78	.58	.59
EIG	2.62							
PC	52.5					Alpha = .76		

The results of the analyses done at the item-level for each subscale for the 12th-grade students are summarized in Table 55. As Table 55 indicates, subscale means ranged from 1.70 for Worry to 3.01 for Effort, and subscale standard deviations ranged from .64 for Worry to .77 for Effort. These results are very similar to the results obtained for 8th-grade students. The subscale mean for 8th-grade students for Worry was 1.75 and for Effort was 3.38 as compared with 1.70 and 3.01 respectively for the 12th-grade students, but the subscale reliabilities for the 12th-grade students were generally higher than those for the 8th-grade students. The Alpha coefficients of the six subscales for 12th-grade students ranged from .73 for Self-checking to .85 for Effort. Tables 56 through 61 summarize the results of analyses performed on item-level data for the 12th-grade subjects of the main study. These tables are comparable with those summarizing item-level analyses which were presented earlier. Comparisons of these results show how elimination of extra items was done and how the removal of some of poor items affected the reliability of the subscales.

Table 55

Number of Items, Mean, Standard Deviation and Cronbach's Alpha for the Main Study 12th Grade

Variable	# of Items	Mean	SD	Alpha
AWARE	5	2.84	.70	.78
COGSTR	5	2.66	.73	.77
PLAN	5	2.76	.72	.78
SELFCHK	5	2.52	.68	.73
WORRY	8	1.70	.64	.83
EFFORT	5	3.01	.77	.85

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort.

Table 56

SUBSCALE: Awareness for Main Study 12th. Grade Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
28	.69				3.22	.91	.51	.48
40	.76				2.71	.98	.59	.57
63	.68				2.47	.99	.50	.47
53	.78				2.86	.94	.61	.60
48	.74				2.96	.97	.57	.55
EIG	2.67							
PC	53.4					Alpha = .78		

Table 57

SUBSCALE: Cognitive Strategy for Main Study 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
34	.72				2.74	1.05	.54	.51
43	.66				2.70	1.03	.48	.44
47	.73				2.66	1.02	.56	.54
55	.73				2.61	.98	.55	.54
66	.77				2.61	.97	.60	.60
EIG	2.62							
PC	52.5					Alpha = .77		

Table 58

SUBSCALE: Planning for Main Study 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
39	.66				2.68	1.06	.49	.43
**[a]	.67				2.38	1.01	.50	.44
**[a]	.80				2.93	.94	.64	.65
49	.75				2.72	.96	.56	.56
58	.79				3.06	.95	.61	.62
EIG	2.70							
PC	54.0					Alpha = .78		

[a] ** = Denotes new items that were initially introduced with the High School Students Prior to the Pilot Study and that were carried over to the Main Study, 12th Grade.

Table 59

SUBSCALE: Self-checking for Main Study 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
@[a]	.76				2.68	1.01	.54	.58
35	.74				2.68	.96	.53	.55
@[a]	.47				2.38	1.02	.30	.22
46	.66				2.34	.98	.46	.44
**[b]	.81				2.53	.98	.62	.66
EIG	2.44							
PC	48.8					Alpha = .73		

[a] @ = Denotes new items that were introduced for the Main Study, 12th Grade.

[b] ** = Denotes new items that were initially introduced with the High School Students Prior to the Pilot Study and that were carried over to the Main Study, 12th Grade.

Table 60

SUBSCALE: Worry for Main Study 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
20	.58				1.62	.89	.46	.34
82	.73				1.49	.90	.61	.54
89	.66				1.34	.80	.53	.43
80	.77				1.61	.93	.65	.59
87	.71				1.50	.83	.59	.50
95	.62				2.02	1.00	.51	.39
79	.68				2.02	1.12	.57	.47
93	.71				1.99	1.01	.61	.51
EIG	3.76							
PC	47.0					Alpha = .83		

Table 61

SUBSCALE: Effort for Main Study 12th Grade. Item Number, Mean, Standard Deviation, Item-total Correlation, Communalities and Cronbach's Alpha for the Main Study 12th Grade

	Factor Loadings							
Item#	F1	F2	F3	F4	Mean	SD	R(IT)	COMM
15	.79				2.97	.93	.66	.63
30	.86				2.88	.97	.75	.74
10	.85				2.92	1.01	.73	.72
45	.60				3.09	.93	.45	.36
27	.85				3.20	.98	.74	.73
EIG	3.17							
PC	63.6					Alpha = .85		

Finally, Table 62 compares the last reduced version of the instrument (33-item form) with the original 100-item version (the Curiosity subscale is not included in the final version). It should be noted that items were added to the original 100-item pool. We compare the original version with the final version in number of items, number of factors and the size of Alpha. As Table 62 indicates, the number of items for some of the subscales was reduced dramatically. For example, the Effort subscale in the original form had 31 items and was reduced to only 5 items in the final version. Awareness had 8 items and was reduced to 5, Cognitive Strategy was reduced from 14 to 5 (but new items were added), Planning from 9 to 5 (but new items were added), Self-checking from 8 to 5, and Worry from 14 to 8. The number of factors for the subscales in the original version ranged from 2 to 7 factors, There was not one subscale in the original form within which all items load on one factor, that is, items of none of the subscales in the original form fell under a single category. In the final version instrument, however, all items within any of the six subscales loaded on only one factor, which means that under each category there was only one category on one dimension of items. In other words, in the final version we have more homogeneous sets of items under the subscales than in the original form. The Alpha coefficients of the subscales of the original and the final versions were very close. Reduction of items did not have much effect on the reliabilities of the subscales. For example, the most interesting part of this table is the comparison of the Effort subscale in the full and reduced form. The Effort subscale in the original form had 31 items with Alpha of .84. In the final version, this subscale had only 5 items and the Alpha was .85.

As indicated earlier, comparing the original form with the reduced form on two different groups of subjects may not be a valid comparison; however, comparable results of the two forms (original and final versions) obtained from two different groups indicate that, in a sense, the scales were cross-validated.

As mentioned earlier, principal components analysis was performed on the items within each subscale to see if items were unidimensional within a subscale. Normally, a confirmatory factor analysis should follow exploratory analysis to see if the selected items fit under a specific subscale. Confirmatory factor analysis, however, was not done because of the limitation of number of subjects within any single study group. Combining different groups of subjects

Table 62

Number of Items, Number of Factors and Alpha Coefficients for the Full 100-item State Instrument and the Reduced 12th Grade Main Study

	Number of Items		Number of Factors		Alpha	
Subscale	Full	Reduced	Full	Reduced	Full	Reduced
AWARE	8	5	2	1	.78	.78
COGSTR	14	5	4	1	.81	.77
CURIOS*	10	N/A	2	N/A	.84	N/A
PLAN	9	5	2	1	.80	.78
SELFCHK	8	5	2	1	.77	.73
WORRY	14	8	3	1	.90	.83
EFFORT	31	5	7	1	.84	.85

Note. AWARE = Awareness; COGSTR = Cognitive Strategy; CURIOS = Curiosity; PLAN = Planning; SELFCHK = Self-checking; WORRY = Worry; EFFORT = Effort.

[a] Curiosity was not included in the Reduced Version of the 12th Grade Main Study.

on whom the metacognitive instrument was applied could give enough subjects to satisfy the confirmatory analysis subject requirement, but the problem in combining the groups is the lack of exact comparability of metacognitive items across the groups of subjects.

References

Edwards, A.L. (1961). *Statistical methods for the behavioral sciences*. New York: Holt, Rinehart and Winston.

Khabiri, P. (1993). *The role of metacognition, effort and worry in math problem solving requiring problem translation*. Unpublished doctoral dissertation. University of Southern California, Los Angeles.

Kosmicki, J. (1993). *The effect of differential test instructions on math achievement, effort, and worry of community college students*. Unpublished doctoral dissertation. University of Southern California, Los Angeles.

Morris, L.W., Davis, M.A., & Hutchings, C.H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology*, *73*(4), 541-555.

O'Neil, H.F., Jr., Baker, E.L., Jacoby, A., Ni, Y., & Wittrock, M. (1990). *Human benchmarking studies of expert systems*. Los Angeles: University of California, Center for Technology Assessment.

O'Neil, H.F., Baker, E.L., & Matsuura, S. (1992). Reliability and validity of Japanese children's trait and state worry and emotionality scales. *Anxiety, Stress, and Coping*, *5*, 225-239.

Spielberger, C.D., Peters, R.A., & Frain, F.J. (1976). *The State-Trait Curiosity Inventory* (unpublished test). Tampa: University of South Florida.

Spielberger, C.D., Peters, R.A., & Frain, F.J. (1981). Curiosity and anxiety. In H.G. Voss & H. Keller (Eds.), *Curiosity research: Basic concepts and results*. Weinheim, Germany: Beltz.

APPENDIX B: Administration Script – Main Study

SESSION ADMINISTRATION SCRIPT

[NOTE: INSTRUCTIONS TO THE ADMINISTRATOR ARE IN BOLD CAPITAL LETTERS AND SHOULD NOT BE READ TO THE STUDENTS.]

INTRODUCTION

Hello. My name is _____. Today you will be participating in a nationwide study of students your age. To make sure that all students receive the same instructions, I will be reading them to you from a script.

This study is the National Assessment of Educational Progress. Its purpose is to provide information on the knowledge and attitudes of young people throughout the United States. As part of this study, you will answer questions about yourself and about mathematics. It will take about 45 minutes. You will not be allowed to ask questions during the assessment.

By doing the best you can, you will be making an important contribution.

DISTRIBUTING BOOKLET

Before I hand out your materials, please clear your desks. As I call your name, please raise your hand and I will put an envelope and pencil on your desk. Do not take the test booklet out of the envelope yet.

DISTRIBUTE THE ENVELOPES AND PENCILS TO THE STUDENTS. (ENLIST THE HELP OF THE TEACHER AND/OR A FEW CAPABLE STUDENTS TO HELP YOU DISTRIBUTE THE ENVELOPES AND PENCILS AS QUICKLY AS POSSIBLE.)

ASK IF ANY STUDENT HAS NOT RECEIVED AN ENVELOPE. GIVE THOSE STUDENTS SUPPLEMENTARY ENVELOPES AND ASK THEM TO WRITE THEIR NAME ON THE LABEL.

WHEN ALL STUDENTS HAVE ENVELOPES AND PENCILS, PROCEED AS FOLLOWS:

Open your envelope and take out the booklet. Turn the booklet face down on your desk.

CHECK THAT ALL STUDENTS HAVE TAKEN THE BOOKLET OUT.

CODING THE BOOKLET

Please turn your booklet over. Code your grade, birth date, and sex in the box in the middle of the page. Write the number of your grade in the box labeled "Grade." Then fill in the oval next to the number in the grid below the box. In the boxes labeled "Birthday," write the month and year you were born and fill in the correct ovals. Next, write "M" for male or "F" for female in the box labeled "Sex" and fill in the correct oval. Be sure to fill in the ovals completely.

BOOKLET DIRECTIONS

Now open your booklet to the Directions on the first page. Read them to yourself as I read them out loud.

This assessment uses many different booklets each with different questions. Do not worry if the person next to you is working on questions that do not look like those you are working on.

Read each question carefully and answer it as well as you can. Do not spend too much time on any one question.

Each booklet has three parts. We will do the four sample questions together and you will complete the other parts on your own. You will be told when to begin each part. Stop when you see this sign.

If you finish a part early, you may check your work on that part only. Do not begin another part until you are told to continue.

Now read sample 1. The choices for some questions will be written across the page as shown. Fill in the oval for the best answer. **READ SAMPLE 1 AND ANSWER CHOICES.**

SAMPLE 1	Almost every day	Once or twice a week	Once or twice a month	Never or hardly ever
1. How often do you watch movies on TV?	(A)	(B)	(C)	(D)

There is no best answer to this question. Your answer will tell us how often you watch movies on TV.

Now read sample 2. Fill in the oval for the choice that you think is correct. **READ SAMPLE 2 AND ANSWER CHOICES.**

SAMPLE 2

2. How many minutes are there in 2 hours?

   (A) 12

   (B) 24

   (C) 60

   (D) 120

You should have filled in the oval for "120" because there are 120 minutes in 2 hours.

Now read sample 3 and write your answer on the blank line below. **READ SAMPLE 3.**

SAMPLE 3

3. What kind of music do you like best?

   (Write in) _____

You should answer this question by writing the kind of music you like best. Sometimes there will be more than one line on which to write your answer. Use as many lines as you need for your answer.

Now read sample 4. For some of the questions you may need to write or draw the answer. You can see how this is done in sample 4. **READ SAMPLE 4.**

SAMPLE 4

4. Draw a triangle in the space below.

Remember:

- Read each question CAREFULLY.
- Fill in only ONE OVAL for each question or write your answer in the space provided.
- If you change your answer, ERASE your first answer COMPLETELY.
- CHECK OVER your work if you finish a section early.

Now put your pencils down while I read the instructions for the assessment.

BOOKLET SECTIONS

We are ready to begin the assessment now. I cannot answer any questions during the assessment. If you have a question, save it until the end of the class and I will answer questions then. If you need another pencil at any time, raise your hand and I will bring one to you. If you need to do some calculations to get an answer, do them in the booklet.

Turn to the orange page—where the Directions for Sections 1 and 2 begin.

TIMING BOOKLET SECTIONS

SECTION 1: Read the directions for Sections 1 and 2. Look up at me when you have finished reading. **WAIT NO MORE THAN 45 SECONDS.** Now turn the page to the beginning of Section 1. You will have 15 minutes for Section 1. **NOTE THE TIME ON YOUR WATCH AND CALCULATE WHEN 15 MINUTES WILL HAVE ELAPSED.**

SAY: Please begin.

AFTER 15 MINUTES, SAY: Please stop.

SECTION 2: Now turn the page to the first yellow page where the Directions for Sections 1 and 2 are repeated. Read the directions for Sections 1 and 2 again. Look up when you have finished reading. **WAIT NO MORE THAN 45 SECONDS.**

Now turn the page to the beginning of Section 2. You will have 15 minutes for Section 2. **NOTE THE TIME ON YOUR WATCH AND CALCULATE WHEN 15 MINUTES WILL HAVE ELAPSED.**

SAY: Please begin.

AFTER 15 MINUTES SAY: Please stop.

SECTION 3: (Self-Assessment Measure) Now turn the page to the beginning of Section 3, the first blue page. You will have 10 minutes to read the instructions and complete the items in Section 3. Be sure to read the instructions before you begin answering the questions. **NOTE THE TIME ON YOUR WATCH AND CALCULATE WHEN 10 MINUTES WILL HAVE ELAPSED. IF THERE ARE LESS THAN 10 MINUTES LEFT IN THE PERIOD, THEN REDUCE THE 10 MINUTES TO WHATEVER TIME IS LEFT. (YOU WILL NEED TO LEAVE A COUPLE OF MINUTES AT THE END TO PICK UP THE ENVELOPES.)**

SAY: Please begin.

AFTER 10 MINUTES, SAY: Please stop working and close your booklets.

RETURN OF BOOKLETS TO ENVELOPES

Put your booklet back in the envelope. Fasten the envelope. Do not lick it.

ENDING THE SESSION AND PICKING UP ENVELOPES AND PENCILS

Before I pick up your envelopes and pencils, I would like to thank you for being part of our study. We'll be sending each of you a letter next month which will contain your results as well as anything else we promised you in the directions you read.

PICK UP THE ENVELOPES AND PENCILS.

TURN STUDENTS OVER TO THEIR TEACHER OR TELL THEM TO GO TO THEIR NEXT CLASS.

APPENDIX C: Tables of ANOVA Results

Table A1

Financial Incentives Pilot Study 1, Grade 8: Summary of Analysis of Variance on Total Mathematics Score (N=158)

Source	SS	df	MS	F	Prob F
Treatment	208.8	3	69.6	1.8	.147
Ethnicity	1355.1	2	677.5	17.7	.001
Gender	53.0	1	53.0	1.4	.242
Treatment x Ethnicity	191.0	6	31.8	.8	.549
Treatment x Gender	98.5	3	32.8	.9	.466
Ethnicity x Gender	8.9	2	4.4	.1	.891
Treatment x Ethnicity x Gender	221.7	6	37.0	1.0	.452
Residual	5136.7	134	38.3		
Total	7476.2	157	47.6		

Table A2

Financial Incentives Pilot Study 1, Grade 8: Summary of Analysis of Variance on Moderately Difficult Mathematics Items (N=158)

Source	SS	df	MS	F	Prob F
Treatment	66.7	3	22.2	3.8	.012
Ethnicity	235.4	2	117.7	20.1	.001
Gender	9.2	1	9.2	1.6	.213
Treatment x Ethnicity	50.4	6	8.4	1.4	.207
Treatment x Gender	18.2	3	6.1	1.0	.380
Ethnicity x Gender	6.9	2	3.4	.6	.556
Treatment x Ethnicity x Gender	29.8	6	5.0	.8	.537
Residual	785.8	134	5.9		
Total	1220.3	157			

Table A3

Financial Incentives Pilot Study 1, Grade 8: Summary of Analysis of Variance on Number of Mathematics Items Omitted (N=158)

Source	SS	df	MS	F	Prob F
Treatment	43.1	3	14.4	8.7	.001
Ethnicity	12.9	2	6.4	3.9	.022
Gender	.3	1	.3	.2	.676
Treatment x Ethnicity	59.4	6	9.9	6.0	.001
Treatment x Gender	2.3	3	.8	.5	.702
Ethnicity x Gender	5.4	2	2.7	1.6	.197
Treatment x Ethnicity x Gender	10.8	6	1.8	1.1	.368
Residual	220.5	134	1.6		
Total	334.9	157	2.1		

Table A4

Financial Incentives Pilot Study 1, Grade 8: Summary of Analysis of Variance on Number of Mathematics Items Not Attempted (N=158)

Source	SS	df	MS	F	Prob F
Treatment	150.7	3	50.2	6.2	.001
Ethnicity	6.0	2	3.0	.4	.691
Gender	13.3	1	13.3	1.7	.200
Treatment x Ethnicity	145.0	6	24.2	3.0	.009
Treatment x Gender	1.3	3	.4	.1	.983
Ethnicity x Gender	25.2	2	12.6	1.6	.213
Treatment x Ethnicity x Gender	31.7	6	5.3	.7	.685
Residual	1080.0	134	8.1		
Total	1402.7	157	8.9		

Table A5

Financial Incentives Pilot Study 1, Grade 8: Summary of Analysis of Variance on Self-checking (N=157)

Source	SS	df	MS	F	Prob F
Treatment	3.6	3	1.2	2.9	.039
Ethnicity	1.2	2	.6	1.5	.229
Gender	.14	1	.1	.3	.560
Treatment x Ethnicity	5.8	6	1.0	2.3	.035
Treatment x Gender	.3	3	.1	.2	.867
Ethnicity x Gender	.9	2	.4	1.1	.347
Treatment x Ethnicity x Gender	2.7	6	.5	1.1	.368
Residual	55.0	133	.4		
Total	68.1	156	.4		

Table A6

Financial Incentives Pilot Study 1, Grade 8: Summary of Analysis of Variance on Effort (N=156)

Source	SS	df	MS	F	Prob F
Treatment	.4	3	.1	.3	.802
Ethnicity	3.5	2	1.7	4.0	.020
Gender	2.3	1	2.3	5.4	.002
Treatment x Ethnicity	1.4	6	.2	.5	.408
Treatment x Gender	1.9	3	.6	1.5	.776
Ethnicity x Gender	.7	2	.3	.8	.225
Treatment x Ethnicity x Gender	2.1	6	.4	.8	.562
Residual	56.9	132	.4		
Total	70.4	155	.5		

Table A7

Financial Incentives Pilot Study 1, Grade 8: Summary of Analysis of Variance on Mathematics Block 3 (N=158)

Source	SS	df	MS	F	Prob F
Treatment	106.7	3	35.5	2.1	.101
Ethnicity	411.7	2	205.8	12.3	.001
Gender	77.0	1	77.0	4.6	.034
Treatment x Ethnicity	89.0	6	14.8	.9	.509
Treatment x Gender	58.4	3	19.5	1.2	.327
Ethnicity x Gender	9.8	2	4.9	.3	.748
Treatment x Ethnicity x Gender	86.9	6	14.5	.9	.524
Residual	2248.8	134	16.8		
Total	3164.4	157	20.2		

Table A8

Financial Incentives Pilot Study 1, Grade 12: Summary of Analysis of Variance on Total Mathematics Score (N=200)

Source	SS	df	MS	F	Prob F
Treatment	79.6	3	26.5	.4	.748
Ethnicity	3013.2	2	1506.6	23.1	.001
Gender	44.8	1	44.8	.7	.408
Treatment x Ethnicity	262.1	6	43.7	.7	.674
Treatment x Gender	214.6	3	71.5	1.1	.351
Ethnicity x Gender	230.0	2	115.0	1.8	.174
Treatment x Ethnicity x Gender	384.6	6	64.1	1.0	.438
Residual	11466.3	176	65.1		
Total	15786.7	199	79.3		

Table A9

Financial Incentives Pilot Study 1, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Not Reached (N=200)

Source	SS	df	MS	F	Prob F
Treatment	5.6	3	1.9	.2	.925
Ethnicity	140.1	2	70.0	5.9	.003
Gender	2.2	1	2.2	.2	.671
Treatment x Ethnicity	80.8	6	13.5	1.1	.346
Treatment x Gender	23.1	3	7.7	.6	.585
Ethnicity x Gender	21.7	2	10.8	.9	.404
Treatment x Ethnicity x Gender	22.8	6	3.8	.3	.926
Residual	2094.3	176	11.9		
Total	2415.0	199	12.1		

Table A10

Financial Incentives Pilot Study 1, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Not Attempted (N=200)

Source	SS	df	MS	F	Prob F
Treatment	2.2	3	.7	.05	.985
Ethnicity	202.2	2	101.1	6.4	.002
Gender	7.1	1	7.1	.5	.503
Treatment x Ethnicity	109.6	6	18.3	1.2	.334
Treatment x Gender	8.3	3	2.8	.2	.913
Ethnicity x Gender	10.0	2	5.0	.3	.731
Treatment x Ethnicity x Gender	35.4	6	5.9	.4	.896
Residual	2786.9	176	15.8		
Total	3196.2	199	16.1		

Table A11

Financial Incentives Pilot Study 1, Grade 12: Summary of Analysis of Variance on Perceived Mathematics Ability (N=136)

Source	SS	df	MS	F	Prob F
Treatment	2.7	3	.9	1.5	.211
Ethnicity	5.1	2	2.5	4.3	.016
Gender	1.6	1	1.6	2.7	.106
Treatment x Ethnicity	1.2	6	.2	.5	.920
Treatment x Gender	1.1	3	.4	.3	.613
Ethnicity x Gender	.9	2	.5	.6	.453
Treatment x Ethnicity x Gender	5.3	6	.9	1.5	.188
Residual	66.3	112	.6		
Total	86.6	135	.6		

Table A12a

Financial Incentives Pilot Study 1, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Omitted (N=200)

Source	SS	df	MS	F	Prob F
Treatment	1.8	3	.6	.2	.893
Ethnicity	6.0	2	3.0	1.0	.368
Gender	17.1	1	17.1	5.8	.017
Treatment x Ethnicity	11.1	6	1.9	.6	.709
Treatment x Gender	7.2	3	2.4	.8	.492
Ethnicity x Gender	5.3	2	2.6	.9	.412
Treatment x Ethnicity x Gender	16.8	6	2.8	.9	.465
Residual	521.9	176	3.0		
Total	591.2	199	3.0		

Table A12b

Financial Incentives Pilot Study 1, Grade 12: Summary of Analysis of Variance on Worry (N=196)

Source	SS	df	MS	F	Prob F
Treatment	2.2	3	.7	1.5	.227
Ethnicity	7.0	2	3.5	7.1	.001
Gender	3.4	1	3.4	6.9	.010
Treatment x Ethnicity	4.4	6	.7	1.4	.194
Treatment x Gender	1.4	3	.5	1.0	.412
Ethnicity x Gender	.6	2	.3	.6	.573
Treatment x Ethnicity x Gender	2.5	6	.4	.8	.548
Residual	85.5	172	.5		
Total	104.2	195	.5		

Table A13

Financial Incentives Pilot Study 2, Grade 12: Summary of Analysis of Variance on Perceived Self-checking (N=170)

Source	SS	df	MS	F	Prob F
Treatment	3.6	3	1.2	3.8	.011
Ethnicity	.1	1	.1	.3	.605
Gender	.1	1	.1	.3	.604
Treatment x Ethnicity	1.9	3	.6	2.0	.118
Treatment x Gender	.3	3	.1	.3	.843
Ethnicity x Gender	.1	1	.6	.2	.672
Treatment x Ethnicity x Gender	.8	3	.3	.9	.451
Residual	47.9	154	.3		
Total	55.2	169	.3		

Table A14

Financial Incentives Pilot Study 2, Grade 12: Summary of Analysis of Variance on Total Mathematics Score (N=170)

Source	SS	df	MS	F	Prob F
Treatment	67.1	3	22.4	.4	.723
Ethnicity	433.9	1	433.9	8.6	.004
Gender	39.8	1	39.8	.8	.376
Treatment x Ethnicity	127.9	3	42.6	.8	.472
Treatment x Gender	92.5	3	30.8	.6	.610
Ethnicity x Gender	58.4	1	58.4	1.2	.284
Treatment x Ethnicity x Gender	125.6	3	41.9	.8	.480
Residual	7788.2	154	50.6		
Total	8818.7	169	52.2		

Table A15

Financial Incentives Pilot Study 2, Grade 12: Summary of Analysis of Variance on Worry (N=169)

Source	SS	df	MS	F	Prob F
Treatment	.8	3	.3	.8	.520
Ethnicity	5.9	1	5.9	16.1	.001
Gender	.1	1	.1	.3	.562
Treatment x Ethnicity	.7	3	.2	.6	.613
Treatment x Gender	.5	3	.2	.5	.694
Ethnicity x Gender	.3	1	.3	.8	.384
Treatment x Ethnicity x Gender	1.0	3	.3	.9	.457
Residual	56.2	153	.4		
Total	65.2	168	.4		

Table A16

Goal Orientation Pilot, Grade 8: Summary of Analysis of Variance on Total Mathematics Score (N=55, students tested first)

Source	SS	df	MS	F	Prob F
Treatment	884.0	3	294.7	3.4	.025
Residual	4436.8	51	87.0		
Total	5320.8	54	98.5		

Table A17

Goal Orientation Pilot, Grade 8: Summary of Analysis of Variance on Total Mathematics Score (N=173)

Source	SS	df	MS	F	Prob F
Treatment	120.0	3	40.0	.7	.580
Ethnicity	1837.9	1	1837.9	30.2	.001
Gender	50.6	1	50.6	.8	.363
Treatment x Ethnicity	77.2	3	25.7	.4	.737
Treatment x Gender	254.3	3	84.9	1.4	.247
Ethnicity x Gender	2.3	1	2.3	.03	.847
Treatment x Ethnicity x Gender	211.1	3	70.4	1.2	.329
Residual	9569.9	157	61.0		
Total	1240.1	172	72.1		

Table A18

Goal Orientation Pilot, Grade 8: Summary of Analysis of Variance on Number of Mathematics Items Not Reached (N=173)

Source	SS	df	MS	F	Prob F
Treatment	109.8	3	36.6	2.4	.074
Ethnicity	23.7	1	23.7	1.5	.218
Gender	99.3	1	99.3	6.4	.012
Treatment x Ethnicity	33.6	3	11.2	.7	.540
Treatment x Gender	101.8	3	33.9	2.2	.092
Ethnicity x Gender	52.3	1	52.3	3.4	.068
Treatment x Ethnicity x Gender	23.5	3	7.8	.5	.679
Residual	2437.3	157	15.5		
Total	2839.9	172	16.5		

Table A19

Goal Orientation Pilot, Grade 8: Summary of Analysis of Variance on Number of Mathematics Items Not Attempted (N=173)

Source	SS	df	MS	F	Prob F
Treatment	127.9	3	42.6	2.4	.073
Ethnicity	49.3	1	49.3	2.7	.100
Gender	100.9	1	100.9	5.6	.019
Treatment x Ethnicity	58.0	3	19.3	1.1	.361
Treatment x Gender	112.4	3	37.5	2.1	.105
Ethnicity x Gender	91.4	1	91.4	5.1	.026
Treatment x Ethnicity x Gender	9.4	3	3.1	.2	.914
Residual	2823.7	157	18.0		
Total	3328.9	172	19.4		

Table A20

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Mathematics Block 3 (N=197)

Source	SS	df	MS	F	Prob F
Treatment	43.8	3	14.6	.9	.431
Ethnicity	579.7	1	579.7	36.6	.001
Gender	96.2	1	96.2	6.1	.015
Treatment x Ethnicity	89.3	3	29.8	1.9	.135
Treatment x Gender	139.9	3	46.6	2.9	.034
Ethnicity x Gender	11.4	1	11.4	.7	.397
Treatment x Ethnicity x Gender	41.8	3	13.9	.9	.452
Residual	2864.1	181	15.8		
Total	4090.1	196	20.9		

Table A21

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Total Mathematics Score (N=197)

Source	SS	df	MS	F	Prob F
Treatment	107.0	3	35.7	.6	.608
Ethnicity	2181.2	1	2181.2	37.4	.001
Gender	320.3	1	320.3	5.5	.020
Treatment x Ethnicity	438.0	3	146.0	2.5	.061
Treatment x Gender	324.4	3	108.1	1.9	.139
Ethnicity x Gender	22.6	1	22.6	.4	.534
Treatment x Ethnicity x Gender	123.2	3	41.1	.7	.550
Residual	10549.6	181	58.3		
Total	14905.0	196	76.0		

Table A22

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Omitted (N=197)

Source	SS	df	MS	F	Prob F
Treatment	11.2	3	3.7	1.5	.214
Ethnicity	22.0	1	22.0	8.9	.003
Gender	4.1	1	4.1	1.6	.202
Treatment x Ethnicity	9.8	3	3.3	1.3	.269
Treatment x Gender	2.1	3	.7	.3	.840
Ethnicity x Gender	1.0	1	1.0	.4	.525
Treatment x Ethnicity x Gender	1.7	3	.6	.2	.877
Residual	447.2	181	2.5		
Total	501.8	196	2.6		

Table A23

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Not Reached (N=197)

Source	SS	df	MS	F	Prob F
Treatment	13.2	3	4.4	.4	.749
Ethnicity	313.2	1	313.2	28.9	.001
Gender	87.6	1	87.6	8.1	.005
Treatment x Ethnicity	60.1	3	20.0	1.9	.139
Treatment x Gender	47.7	3	15.9	1.5	.225
Ethnicity x Gender	58.7	1	58.7	5.4	.021
Treatment x Ethnicity x Gender	28.5	3	9.5	.9	.454
Residual	1958.6	181	10.8		
Total	2632.1	196	13.4		

Table A24

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Not Attempted (N=197)

Source	SS	df	MS	F	Prob F
Treatment	28.9	3	9.6	.7	.574
Ethnicity	501.2	1	501.2	34.6	.001
Gender	129.3	1	129.3	8.9	.003
Treatment x Ethnicity	65.8	3	21.9	1.5	.212
Treatment x Gender	49.9	3	16.6	1.1	.331
Ethnicity x Gender	75.1	1	75.1	5.2	.024
Treatment x Ethnicity x Gender	30.7	3	10.2	.7	.549
Residual	2619.4	181	14.5		
Total	3589.3	196	18.3		

Table A25

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Planning (N=195)

Source	SS	df	MS	F	Prob F
Treatment	1.4	3	.5	1.2	.317
Ethnicity	2.2	1	2.2	5.7	.018
Gender	.03	1	.03	.1	.792
Treatment x Ethnicity	.3	3	.1	.3	.844
Treatment x Gender	.9	3	.3	.8	.484
Ethnicity x Gender	.9	1	.9	2.3	.128
Treatment x Ethnicity x Gender	.3	3	.1	.3	.823
Residual	68.0	179	.4		
Total	74.4	194	.4		

Table A26

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Curiosity (N=195)

Source	SS	df	MS	F	Prob F
Treatment	3.5	3	1.2	2.0	.110
Ethnicity	10.2	1	10.2	17.7	.001
Gender	.04	1	.04	.08	.784
Treatment x Ethnicity	.5	3	.2	.3	.822
Treatment x Gender	.7	3	.2	.4	.729
Ethnicity x Gender	.4	1	.4	.8	.383
Treatment x Ethnicity x Gender	.4	3	.1	.2	.878
Residual	102.9	179	.6		
Total	120.0	194	.6		

Table A27

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Worry (N=195)

Source	SS	df	MS	F	Prob F
Treatment	.5	3	.2	.5	.699
Ethnicity	10.4	1	10.4	29.9	.001
Gender	.3	1	.3	1.0	.328
Treatment x Ethnicity	1.0	3	.3	1.0	.403
Treatment x Gender	2.7	3	.9	2.6	.057
Ethnicity x Gender	.3	1	.3	1.0	.320
Treatment x Ethnicity x Gender	2.4	3	.8	2.3	.080
Residual	62.2	179	14.0		
Total	83.5	194	16.1		

Table A28

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Perceived Mathematics Ability (N=182)

Source	SS	df	MS	F	Prob F
Treatment	1.5	3	.5	.5	.656
Ethnicity	5.0	1	5.0	5.3	.022
Gender	.5	1	.5	.5	.478
Treatment x Ethnicity	.4	3	.1	.2	.929
Treatment x Gender	3.0	3	1.0	1.1	.363
Ethnicity x Gender	.03	1	.03	.03	.869
Treatment x Ethnicity x Gender	2.6	3	.9	.9	.424
Residual	155.1	166	1.5		
Total	172.6	181	1.6		

Table A29

Goal Orientation Pilot, Grade 12: Summary of Analysis of Variance on Perceived Mathematics Grades (N=179)

Source	SS	df	MS	F	Prob F
Treatment	2.2	3	.7	.7	.576
Ethnicity	8.2	1	8.2	7.4	.007
Gender	.4	1	.4	.4	.528
Treatment x Ethnicity	2.0	3	.7	.6	.613
Treatment x Gender	.5	3	.2	.1	.936
Ethnicity x Gender	.4	1	.4	.3	.573
Treatment x Ethnicity x Gender	4.2	3	1.4	1.3	.290
Residual	180.3	163	1.9		
Total	202.7	178	2.1		

Table A30

Main Study, Grade 8: Summary of Analysis of Variance on Easy Mathematics Items (N=749)

Source	SS	df	MS	F	Prob F
Treatment	15.5	3	5.2	2.7	.043
Ethnicity	97.4	3	32.5	17.1	.001
Gender	.4	1	.4	.2	.641
Treatment x Ethnicity	14.4	9	1.6	.8	.573
Treatment x Gender	4.8	3	1.6	.9	.467
Ethnicity x Gender	7.2	3	2.4	1.3	.288
Treatment x Ethnicity x Gender	14.2	9	1.6	.8	.588
Residual	1359.2	717	1.9		
Total	1513.0	748	2.0		

Table A31

Main Study, Grade 8: Summary of Analysis of Variance on Effort (N=745)

Source	SS	df	MS	F	Prob F
Treatment	3.7	3	1.2	3.2	.022
Ethnicity	3.4	3	1.1	2.9	.033
Gender	4.3	1	4.3	11.0	.001
Treatment x Ethnicity	1.1	9	.1	.3	.970
Treatment x Gender	1.2	3	.4	1.1	.360
Ethnicity x Gender	.3	3	.1	.3	.857
Treatment x Ethnicity x Gender	.3	9	.03	.1	1.00
Residual	274.5	713	.39		
Total	288.3	744	.39		

Table A32

Main Study, Grade 8: Summary of Analysis of Variance on Total Mathematics Score (N=749)

Source	SS	df	MS	F	Prob F
Treatment	227.9	3	76.0	1.3	.260
Ethnicity	8436.3	3	2812.1	50.0	.001
Gender	.3	1	.3	.01	.950
Treatment x Ethnicity	411.5	9	45.8	.8	.610
Treatment x Gender	130.3	3	43.4	.8	.513
Ethnicity x Gender	382.1	3	127.4	2.2	.082
Treatment x Ethnicity x Gender	363.3	9	40.4	.7	.698
Residual	40651.0	717	56.7		
Total	50742.3	748	67.9		

Note. Edit based on 9/30 unique ANOVA output.

Table A33

Main Study, Grade 8: Summary of Analysis of Variance on Perceived Mathematics Ability (N=634)

Source	SS	df	MS	F	Prob F
Treatment	.3	3	.1	.2	.926
Ethnicity	18.5	3	6.1	8.4	.001
Gender	4.5	1	4.5	6.2	.013
Treatment x Ethnicity	1.6	9	.2	.2	.987
Treatment x Gender	1.9	3	.6	.9	.463
Ethnicity x Gender	3.4	3	1.1	1.5	.204
Treatment x Ethnicity x Gender	5.4	9	.6	.8	.591
Residual	439.1	602	.7		
Total	475.7	633	.7		

Table A34

Main Study, Grade 8: Summary of Analysis of Variance on Worry (N=745)

Source	SS	df	MS	F	Prob F
Treatment	1.3	3	.4	1.1	.355
Ethnicity	12.5	3	4.2	10.7	.001
Gender	1.3	1	1.3	3.2	.074
Treatment x Ethnicity	4.3	9	.5	1.2	.275
Treatment x Gender	2.3	3	.8	2.0	.114
Ethnicity x Gender	3.0	3	1.0	2.5	.054
Treatment x Ethnicity x Gender	4.4	9	.5	1.2	.269
Residual	279.1	713	.4		
Total	308.2	744	.4		

Table A35

Main Study, Grade 8: Summary of Analysis of Variance on Number of Mathematics Items Not Reached (N=745)

Source	SS	df	MS	F	Prob F
Treatment	1.3	3	.4	.1	.967
Ethnicity	27.8	3	9.3	1.9	.131
Gender	22.4	1	22.4	4.5	.033
Treatment x Ethnicity	18.7	9	2.1	.4	.924
Treatment x Gender	12.1	3	4.0	.8	.484
Ethnicity x Gender	7.0	3	2.3	.5	.701
Treatment x Ethnicity x Gender	63.5	9	7.1	1.4	.170
Residual	3533.8	717	4.9		
Total	3695.0	748	4.9		

Table A36

Main Study, Grade 8: Summary of Analysis of Variance on Self-checking (N=744)

Source	SS	df	MS	F	Prob F
Treatment	1.8	3	.6	1.6	.193
Ethnicity	3.0	3	1.0	2.6	.050
Gender	1.8	1	1.8	4.7	.031
Treatment x Ethnicity	5.5	9	.6	1.6	.111
Treatment x Gender	2.7	3	.9	2.3	.072
Ethnicity x Gender	2.6	3	.9	2.2	.084
Treatment x Ethnicity x Gender	1.4	9	.2	.4	.928
Residual	272.2	712	.4		
Total	289.2	743	.4		

Table A37

Main Study, Grade 8: Summary of Analysis of Variance on Total Mathematics Score (N=444)

Source	SS	df	MS	F	Prob F
Treatment	466.7	3	155.5	3.0	.029
Ethnicity	4275.9	3	1425.3	27.8	.001
Gender	24.3	1	24.3	.5	.491
Treatment x Ethnicity	138.2	9	15.3	.3	.975
Treatment x Gender	51.0	3	17.0	.3	.802
Ethnicity x Gender	219.1	3	73.0	1.4	.235
Treatment x Ethnicity x Gender	161.3	9	17.9	.3	.958
Residual	2118.8	412	51.3		
Total	27008.1	443	61.0		

TableA38

Main Study, Grade 8: Summary of Analysis of Variance on Effort (N=443)

Source	SS	df	MS	F	Prob F
Treatment	3.5	3	1.2	3.7	.012
Ethnicity	.3	3	.1	.3	.837
Gender	2.0	1	2.0	6.2	.013
Treatment x Ethnicity	1.2	9	.1	.4	.924
Treatment x Gender	1.5	3	.5	1.6	.200
Ethnicity x Gender	.9	3	.3	1.0	.386
Treatment x Ethnicity x Gender	1.0	9	.1	.3	.960
Residual	128.8	411	.3		
Total	139.3	442	.3		

Table A39

Main Study, Grade 12: Summary of Analysis of Variance on Total Mathematics Score (N=719)

Source	SS	df	MS	F	Prob F
Treatment	204.3	4	51.1	.9	.470
Ethnicity	13902.6	3	4634.2	80.7	.001
Gender	710.2	1	710.2	12.4	.001
Treatment x Ethnicity	809.3	12	67.4	1.2	.297
Treatment x Gender	130.4	4	32.6	.6	.686
Ethnicity x Gender	103.4	3	34.5	.6	.615
Treatment x Ethnicity x Gender	663.5	12	55.3	1.0	.483
Residual	38990.3	679	57.4		
Total	55673.8	718	77.5		

Table A40

Main Study, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Omitted (N=719)

Source	SS	df	MS	F	Prob F
Treatment	1.0	4	2.5	1.4	.229
Ethnicity	20.3	3	6.8	3.8	.010
Gender	1.8	1	1.8	1.0	.311
Treatment x Ethnicity	33.2	12	2.8	1.6	.098
Treatment x Gender	10.2	4	2.6	1.4	.217
Ethnicity x Gender	1.6	3	.5	.3	.824
Treatment x Ethnicity x Gender	15.8	12	1.3	.7	.707
Residual	1202.1	679	1.8		
Total	1301.3	718	1.8		

Table A41

Main Study, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Not Reached (N=719)

Source	SS	df	MS	F	Prob F
Treatment	21.3	4	5.3	.6	.644
Ethnicity	252.8	3	84.3	9.9	.001
Gender	.5	1	.5	.1	.812
Treatment x Ethnicity	107.6	12	9.0	1.1	.395
Treatment x Gender	43.4	4	10.8	1.3	.278
Ethnicity x Gender	29.1	3	9.7	1.1	.331
Treatment x Ethnicity x Gender	98.3	12	8.2	1.0	.482
Residual	5766.3	679	8.5		
Total	6343.6	718	8.8		

Table A42

Main Study, Grade 12: Summary of Analysis of Variance on Number of Mathematics Items Not Attempted (N=719)

Source	SS	df	MS	F	Prob F
Treatment	20.2	4	5.1	.4	.789
Ethnicity	392.7	3	131.0	11.1	.001
Gender	4.2	1	4.2	.4	.553
Treatment x Ethnicity	130.6	12	10.9	.9	.525
Treatment x Gender	49.5	4	12.4	1.0	.382
Ethnicity x Gender	21.1	3	7.0	.6	.618
Treatment x Ethnicity x Gender	128.6	12	10.7	.9	.540
Residual	8025.9	679	11.8		
Total	8791.0	718	12.2		

Table A43

Main Study, Grade 12: Summary of Analysis of Variance on Self-checking (N=715)

Source	SS	df	MS	F	Prob F
Treatment	2.8	4	.7	1.7	.142
Ethnicity	5.1	3	1.7	4.2	.006
Gender	3.9	1	3.9	9.7	.002
Treatment x Ethnicity	2.9	12	.2	.6	.844
Treatment x Gender	2.7	4	.7	1.7	.154
Ethnicity x Gender	1.2	3	.4	1.0	.413
Treatment x Ethnicity x Gender	3.2	12	.3	.7	.794
Residual	273.9	675	.4		
Total	295.3	714	.4		

Table A44

Main Study, Grade 12: Summary of Analysis of Variance on Worry (N=715)

Source	SS	df	MS	F	Prob F
Treatment	3.5	4	.9	2.4	.051
Ethnicity	14.6	3	4.9	13.1	.001
Gender	.2	1	.2	.6	.421
Treatment x Ethnicity	4.2	12	.3	.9	.503
Treatment x Gender	3.0	4	.7	2.0	.089
Ethnicity x Gender	1.8	3	.6	1.6	.178
Treatment x Ethnicity x Gender	4.9	12	.4	1.1	.361
Residual	250.4	675	.4		
Total	282.8	714	.4		

Table A45

Main Study, Grade 12: Summary of Analysis of Variance on Effort (N=715)

Source	SS	df	MS	F	Prob F
Treatment	2.5	4	.6	1.2	.298
Ethnicity	13.4	3	4.5	8.9	.001
Gender	3.9	1	3.9	7.7	.006
Treatment x Ethnicity	6.1	12	.5	1.0	.431
Treatment x Gender	.8	4	.2	.4	.796
Ethnicity x Gender	.1	3	.03	.1	.979
Treatment x Ethnicity x Gender	3.5	12	.3	.6	.854
Residual	338.4	675	.5		
Total	368.7	714	.5		

Table A46

Main Study, Grade 12: Summary of Analysis of Variance on Perceived Mathematics Ability (N=670)

Source	SS	df	MS	F	Prob F
Treatment	3.9	4	1.0	1.6	.184
Ethnicity	17.5	3	5.8	9.3	.001
Gender	8.5	1	8.5	13.6	.001
Treatment x Ethnicity	6.6	12	.6	.9	.562
Treatment x Gender	3.1	4	.8	1.2	.298
Ethnicity x Gender	4.1	3	1.4	2.2	.089
Treatment x Ethnicity x Gender	3.0	12	.3	.4	.963
Residual	393.6	630	.6		
Total	443.1	669	.7		

APPENDIX D: Text of Test Instructions

EXPERIMENTAL MOTIVATION PILOT STUDIES

TEST INSTRUCTIONS FOR THE FINANCIAL INCENTIVES, GOAL ORIENTATION, AND CONTROL TREATMENTS GRADES 8 AND 12

Attached are the texts of the test instructions that constituted the three financial incentive treatments, the three goal orientation treatments, and the control treatment used in the motivation pilot studies. Please note that we show text of the financial incentive instructions only for Grade 12; the financial incentive instructions for Grade 8 are identical.

FINANCIAL INCENTIVE TREATMENTS

DIRECTIONS FOR SECTIONS 2 AND 3

The next part is a test which is part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both sections of the test include some newly developed items, and some of the items may be difficult. We are giving money to encourage you to try harder and do well on this test.

There are a total of 44 test items in both sections. We will give you 50¢ for each item you answer correctly. For example, if you get 24 items correct, you will get $12.00.

You will get paid after we score the test.

[50 CENTS]

DIRECTIONS FOR SECTIONS 2 AND 3

The next part is a test which is part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both sections of the test include some newly developed items, and some of the items may be difficult. We are giving money to encourage you to try harder and do well on this test.

There are a total of 44 test items in both sections. We will give you $1.00 for each test item you get correct over 8 items.

For example, if you get 24 items correct, you will get $0.00 for the first 8 items and $1.00 for each of the next 16 items. So you would get $16.00 in all.

You will get paid after we score the test.

[$1 AFTER 8]

DIRECTIONS FOR SECTIONS 2 AND 3

The next part is a test which is part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both sections of the test include some newly developed items, and some of the items may be difficult. We are giving money to encourage you to try harder and do well on this test.

There are a total of 44 test items in both sections. We will give each student $16.00 if the class average score is 24 items or more. Thus, if everyone tries harder and answers more items correctly, the class average score will increase. So try hard and see how many items you can answer correctly, so the whole class will benefit.

You will get paid after we score the test.

[CLASS]

GOAL ORIENTATION TREATMENTS

DIRECTIONS FOR SECTIONS 2 AND 3

The next part is a test which is part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both parts of the test include some newly developed items that are meant to be challenging. If you work hard on these items and do well, you should feel a sense of personal accomplishment and feel good about your effort.

We have found that when students think of difficult test items as a challenge, it makes them try harder, have more fun, and perform better. So, if you try to see this test as challenging and try very hard, you will do well.

In brief, concentrate on the test. Try to see it as a challenge and enjoy mastering it.

[TASK]

DIRECTIONS FOR SECTIONS 2 AND 3

The next part is a test which is part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both parts of the test include some newly developed items which have proven to be an accurate measure of mathematical ability. These new test items will allow us to compare your mathematical ability with that of other students in your classroom, in your school, in your school district, and around the world.

How you perform on these test items will tell us something about how good you are at mathematics. The results of our comparing you with others will be reported to you, your school, your teachers, and your parents.

In brief, how you do will tell us how good you are at this kind of test.

[EGO]

DIRECTIONS FOR SECTIONS 2 AND 3

The next part is a test which is part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

It is really important that you do as WELL as you can on this test. The test score you receive will let others see just how well your teachers are doing in teaching you math this year. Your scores will be compared to those of students in other grades here at this school as well as to those of students in other schools in this city. That is why it is extremely important to do the VERY BEST that you can. Do it for YOURSELF, YOUR PARENTS, and YOUR TEACHERS.

[TEACHER]

CONTROL TREATMENT

DIRECTIONS FOR SECTIONS 2 AND 3

The next part is a test which is part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Its purpose is to provide information on the knowledge and attitudes of young people throughout the United States. By doing the best you can, you will be making an important contribution. Because this is a study, your score will not be shown to anyone in the school.

[CONTROL]

MOTIVATION MAIN STUDY

EXPERIMENTAL AND CONTROL TREATMENTS

GRADES 8 AND 12

Attached are the texts of the test instructions that constituted the experimental and control treatments used in the motivation main study. Please note that we show the financial incentive instructions only for Grade 12; the wording of the financial incentive instructions for Grade 8 is identical.

FINANCIAL INCENTIVE TREATMENT

DIRECTIONS FOR SECTIONS 1 AND 2

The next part is a test which was part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both sections of the test include some newly developed items, and some of the items may be difficult. We are giving money to encourage you to try harder and do well on this test.

There are a total of 44 test items in both sections. We will give you $1.00 for each item you answer correctly. For example, if you get 24 items correct, you will get $24.00. If you answer all of the items correctly, you will get $44.00.

You will get paid about three weeks from now, after we score the test. You will receive cash and it will be given to you here at your school.

[$1 PER ITEM CORRECT]

GOAL ORIENTED TREATMENTS

DIRECTIONS FOR SECTIONS 1 AND 2

The next part is a test which was part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both parts of the test include some newly developed items that are meant to be challenging. If you work hard on these items and do well, you should feel a sense of personal accomplishment and feel good about your effort.

We have found that when students think of difficult test items as a challenge, it makes them try harder, have more fun, and perform better. So, if you try to see this test as challenging and try very hard, you will do well.

In brief, concentrate on the test. Try to see it as a challenge and enjoy mastering it.

[TASK]

DIRECTIONS FOR SECTIONS 1 AND 2

The next part is a test which was part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both parts of the test include some newly developed items which are an accurate measure of mathematical ability. These new test items will allow us to compare your mathematical ability with that of other students in your classroom, in your school, in your school district, and around the world.

How you perform on these test items will tell us something about how good you are at mathematics. The results of our comparing you with others will be reported to you, your school, your teachers, and your parents.

In brief how you do will tell us how good you are at this kind of test.

[EGO]

CERTIFICATE TREATMENT

DIRECTIONS FOR SECTIONS 1 AND 2

The next part is a test which was part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Both parts of the test include some newly developed items which have proven to be an accurate measure of mathematical ability. These new test items will allow us to compare your mathematical ability with that of other students in your classroom, in your school, in your school district, and around the world.

We will provide a UCLA certificate of accomplishment to the students in your class who score in the top 10% on this math test. The certificates could be used to demonstrate your math achievement at job interviews or in the college application process.

We will provide the certificates in about three weeks, after we have scored the tests. You will be given the certificates here at your school.

[CERTIFICATE]

CONTROL TREATMENT

DIRECTIONS FOR SECTIONS 1 AND 2

The next part is a test which was part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each.

Its purpose is to provide information on the knowledge and attitudes of young people throughout the United States. By doing the best you can, you will be making an important contribution. Because this is a study, your score will not be shown to anyone in the school.

[CONTROL]

APPENDIX E: Metacognitive Measure – Main Study, Grade 12

Section 3

Self-Assessment Questionnaire (S12)

Directions: A number of statements which people have used to describe themselves are given below. Read each statement and indicate how you thought or felt during the test. Find the word or phrase which best describes how you thought or felt and circle the number for your answer. There are no right or wrong answers. Do not spend too much time on any one statement. Remember, give the answer which seems to describe how you thought or felt during the test.

	Not at All	Somewhat	Moderately So	Very Much So
1. I was afraid that I should have studied more for this test.	1	2	3	4
2. I concentrated fully when taking the test.	1	2	3	4
3. I was aware of my own thinking.	1	2	3	4
4. I checked my work while I was doing it.	1	2	3	4
5. I attempted to discover the main ideas in the test questions.	1	2	3	4
6. I tried to understand the goals of the test questions before I attempted to answer.	1	2	3	4
7. I felt that others would be disappointed in me.	1	2	3	4
8. I worked as hard as possible.	1	2	3	4
9. I was aware of which thinking technique or strategy to use and when to use it.	1	2	3	4
10. I thought everybody else studied more than I.	1	2	3	4
11. I corrected my errors.	1	2	3	4
12. I asked myself how the test questions related to what I already knew.	1	2	3	4
13. I tried to determine what the test required.	1	2	3	4
14. I thought my score was bad, so everybody including myself would be disappointed.	1	2	3	4
15. I put forth my best effort.	1	2	3	4
16. I was aware of the need to plan my course of action.	1	2	3	4

GO ON TO THE NEXT PAGE

Section 3

	Not at All	Somewhat	Moderately So	Very Much So
17. I almost always knew how much of the test I had left to complete.	1	2	3	4
18. I thought through the meaning of the test questions before I began to answer them.	1	2	3	4
19. I made sure I understood just what had to be done and how to do it.	1	2	3	4
20. I felt regretful.	1	2	3	4
21. I kept working, even on difficult test questions.	1	2	3	4
22. I was aware of my ongoing thinking processes.	1	2	3	4
23. I wasn't happy with my performance.	1	2	3	4
24. I kept track of my progress and, if necessary, I changed my techniques or strategies.	1	2	3	4
25. I used multiple thinking techniques or strategies to solve the test questions.	1	2	3	4
26. I determined how to solve the test questions.	1	2	3	4
27. I was concerned about what would happen if I did poorly.	1	2	3	4
28. I tried to do my best on the test.	1	2	3	4
29. I was aware of my trying to understand the test questions before I attempted to solve them.	1	2	3	4
30. I checked my accuracy as I progressed through the test.	1	2	3	4
31. I selected and organized relevant information to solve the test questions.	1	2	3	4
32. I tried to understand the test questions before I attempted to solve them.	1	2	3	4
33. I did not feel very confident about my performance on this test.	1	2	3	4

GO ON TO THE NEXT PAGE

34. As we mentioned in the directions, we used many booklets each with different questions. We are interested in how well you remember the directions that were given. The directions began with the following statement:

"The next part is a test which was part of the National Assessment of Educational Progress. It contains two sections of 15 minutes each."

Your directions were (choose one):

Ⓐ "Both sections of the test include newly developed items that are meant to be challenging. . . . In brief, concentrate on the test. Try to see it as a challenge and enjoy mastering it."

Ⓑ "These new test items will allow us to compare your mathematical ability with that of other students in your classroom, in your school, in your school district, and around the world. . . . In brief, how you do will tell us how good you are at this kind of test."

Ⓒ "By doing the best you can, you will be making an important contribution. Because this is a study, your score will not be shown to anyone in the school."

Ⓓ "Both sections of the test include some newly developed items, and some of the items may be difficult. We are giving money to encourage you to try harder and do well on this test."

Ⓔ "We will provide a UCLA certificate of accomplishment to the students in your class who score in the top 10% on this math test. The certificates could be used to demonstrate your math achievement at job interviews or in the college application process."

Ⓕ I can't remember the directions.

We are also interested in your assessment of your math ability. Please fill in the oval for your answer to the following question:

35. Compared to your classmates, your math ability is:

Ⓐ High (much better than most of my classmates)
Ⓑ Above average (better than most of my classmates)
Ⓒ Average (equal to most of my classmates)
Ⓓ Below average (less than most of my classmates)
Ⓔ Low (much less than most of my classmates)

We thank you for your participation. We will provide feedback after we score the various tests. Again, thanks.

NAEP MAIN TEST SCORING SCALES May-June 1992

STATE POST THINKING QUESTIONNAIRE Grade 12

Scales	Items
AW=Awareness	3, 9, 16, 22, 29
CS=Cognitive Strategy	5, 12, 18, 25, 31
P=Planning	6, 13, 19, 26, 32
SC=Self-Checking	4, 11, 17, 24, 30
W=Worry	1, 7, 10, 14, 20, 23, 27, 33
EF=Effort	2, 8, 15, 21, 28

Metacognitive = AW + CS + P + SC

AWARENESS

3. I was aware of my own thinking. .....AW
9. I was aware of which thinking technique or strategy to use and when to use it. .....AW
16. I was aware of the need to plan my course of action. .....AW
22. I was aware of my ongoing thinking processes. .....AW
29. I was aware of my trying to understand the test questions before I attempted to solve them. .....AW

COGNITIVE STRATEGY

5. I attempted to discover the main ideas in the test questions. .....CS
12. I asked myself how the test questions related to what I already knew. .....CS
18. I thought through the meaning of the test questions before I began to answer them. .....CS
25. I used multiple thinking techniques or strategies to solve the test questions. .....CS
31. I selected and organized relevant information to solve the test questions. .....CS

PLANNING

6. I tried to understand the goals of the test questions before I attempted to answer. .....P
13. I tried to determine what the test required. .....P
19. I made sure I understood just what had to be done and how to do it .....P
26. I determined how to solve the test questions. .....P
32. I tried to understand the test questions before I attempted to solve them. .....P

SELF-CHECKING

4. I checked my work while I was doing it. .....SC
11. I corrected my errors. .....SC
17. I almost always knew how much of the test I had left to complete. .....SC
24. I kept track of my progress and, if necessary, I changed my techniques or strategies. .....SC
30. I checked my accuracy as I progressed through the test. .....SC

WORRY

1. I was afraid that I should have studied more for this test. .....W
7. I felt that others would be disappointed in me. .....W
10. I thought everybody else studied more than I. .....W
14. I thought my score was bad, so everybody including myself would be disappointed. .....W
20. I felt regretful. .....W
23. I wasn't happy with my performance. .....W
27. I was concerned about what would happen if I did poorly. .....W
33. I did not feel very confident about my performance on this test. .....W

EFFORT

2. I concentrated fully when taking the test. .....EF
8. I worked as hard as possible. .....EF
15. I put forth my best effort. .....EF
21. I kept working, even on difficult test questions. .....EF
28. I tried to do my best on the test. .....EF

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").