MOTIVATION TO DO WELL ON LOW-STAKES TESTS

A Dissertation presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

JAMES S. COLE

Thesis supervisor: Dr. David A. Bergin

MAY 11, 2007

The undersigned, appointed by the dean of the Graduate School,

have examined the dissertation entitled

MOTIVATION TO DO WELL ON LOW-STAKES TESTS

presented by James S. Cole,

a candidate for the degree of doctor of philosophy of educational psychology,

and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. David Bergin

_____

Dr. Steve Osterlind

_____

Dr. Jessica Summers

_____

Dr. Stephen Whitney

_____

Dr. Kennon Sheldon

DEDICATION

To advance this far in my education would not have happened without the endless support of my parents for all these years, especially during my teenage years when I surely pushed them to their limits. Thank you Mom and Dad!

DEDICATED TO CHARLES AND FAY COLE

ACKNOWLEDGEMENTS

I want to thank Dave Bergin for his support, advice, research guidance, and keeping me on track. I also want to thank the rest of my committee which includes Steve Osterlind, Jessica Summers, Ken Sheldon, and Stephen Whitney for their guidance and support. Your feedback and support not only improves the quality of this dissertation, but it also improves my skills as a researcher. Thanks!

I am compelled to also thank Gypsy Denzine. It is because of her that I even pursued a degree in educational psychology. I count myself fortunate that I had the opportunity to work with Gypsy on many research projects, presentations, and papers. Thanks Gypsy.

I must also thank all the people at ARC. Roberta Mullen, Tim Parshall, Robin Albee, Rita Holly, Pam Humphreys, and others at ARC not only encouraged me to pursue my degree but also provided me the flexibility to do so. Thank you ARC.

Finally, I want to acknowledge the support and encouragement of my wife, Shu. It took me awhile, but she always knew I would finish. Thanks!

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF FIGURES vi

LIST OF TABLES vii

ABSTRACT viii

Chapter

1. INTRODUCTION 1

    Purpose of the Study 1

    Prevalence and Use of Low-Stakes Tests 2

    Test Consequences 3

2. LITERATURE REVIEW 6

    Test-Taking Motivation 6

    Self-Determination Theory 10

    Social Context and Low Interest Tasks 18

    Rewards 22

    Conclusion 29

    Hypotheses 30

3. METHOD 31

    Pilot Study 31

        Method and Findings 31

        Participants 31

        Procedure and Experimental Conditions 31

        Findings 34

Main Study 37
Power Analysis 37
Recruitment Procedures 37
Participants 38
Procedure 38
Study Variables 42
Math Test Score 42
Test-Taking Effort 42
Interest, Value, and Perceived Choice 43
Autonomy 44
Gender and ACT Scores 46
Analysis of Data 47
4. RESULTS 49
Preliminary Analyses 49
Effect of Rationale, Choice, and Usefulness and Reward on Internalization, Autonomy, Effort, and Number of Test Items Completed 49
Effect of Rationale, Choice, and Usefulness and Reward on the Math Achievement Test 52
There will be no difference between the three groups regarding the impact of value/usefulness, choice, interest, and effort on test performance 55
Autonomy Group Results 56
Control Group Results 57
Lottery Group Results 58
Summary of Path Analysis 59
5. DISCUSSION 61

Effectiveness of Autonomy Support and Financial Reward in Enhancing Test-Taking Motivation....62

Effectiveness of Autonomy Support and Financial Reward in Improving Test Performance....63

Impact of Autonomy Support and Lottery on Theoretical Understanding of Test-Taking Motivation....65

Implications....67

Limitations....70

Future Research....72

REFERENCES....74

APPENDICES

1. APPENDIX A: Informed Consent....82
2. APPENDIX B: Survey....84
3. APPENDIX C: Demographic Characteristics....89
4. APPENDIX D: Factor Loadings....91
5. APPENDIX E: Intercorrelations....94
6. APPENDIX F: Path Model Results (unstandardized)....97

VITA....100

LIST OF FIGURES

Figure Page

1. The Self-Determination Continuum 14
2. Initial Path Model 48
3. Gender Interaction with Treatment 54
4. Final Path Model Results for Autonomy Group (standardized values) 57
5. Final Path Model Results for Control Group (standardized values) 58
6. Final Path Model Results for Lottery Group (standardized values) 59

LIST OF TABLES

Table Page

1. Means and Standard Deviations....50
2. Scheffes' Post-Hoc Test Results....51
3. Adjusted Mean Test Performance Scores by Group....52
4. Adjusted Mean Test Performance Scores by Gender....53
5. Adjusted Mean Test Performance Scores by Group and Gender....54
6. Summary of model fit statistics....56

ABSTRACT

According to the self-determination theory, students provided with choice, a rationale, and acknowledgement should report higher levels of motivation and test performance on a low-stakes exam compared to other students. However, there is also some evidence that financial, performance-contingent incentives may facilitate students' test-taking motivation and performance. This experimental study investigated differences in test-taking motivation and test score on a math achievement test due to three interventions: autonomy-support, lottery, or control. The autonomy support group received instructions that included a meaningful rationale, acknowledgment that the task might not be interesting, and avoidance of controlling language. The lottery group received 0, 1, or 2 chances in the lottery, depending on their test scores.

For the autonomy and control groups, test taking effort had a significant impact on test performance. However, students who received autonomy support reported higher levels of test effort, but scored significantly lower than students in the lottery group. The performance-based lottery system improved test scores, but only for male test-takers.

Based on the results it appears that the two interventions activate different relationships of key variables during a testing situation. The lottery intervention significantly favored males during the math exam. On the other hand, the autonomy intervention minimized gender differences, and facilitated the role of interest and effort during the exam.

Motivation to do well on low-stakes tests

CHAPTER I: INTRODUCTION

Purpose of the Study

Throughout the school year, students participate in a number of academic activities, some of which students enjoy and some they do not (Baines & Stanely, 2003). One situation that many students may not enjoy is taking standardized achievement tests. This may be especially true for standardized exams that students are required to take, but which have no meaningful outcome or consequence to the student (Cole & Bergin, 2006; Paris, Lawton, Turner, & Roth, 1991; Smith & Smith, 2004). Often referred to as "low stakes" exams, the concern is that students in these testing situations may not be highly motivated to try their best (Smith & Smith, 2005; Sundre & Kitsantas, 2004; Wise & DeMars, 2005). Recognizing that test performance is a function of both knowledge and motivation, the possibility of low student motivation raises the concern of whether data collected are a valid measure of student achievement (Eklof, 2006; Wainer, 1993). As noted by Eklof (2006), "ignoring the test-taking motivation component in low-stakes achievement testing could lead to a confounding of knowledge and motivation and thereby be a threat to the validity of the results" (p. 644). This concern has proven to be a major challenge for educational institutions that make decisions based on these test results. Institutions want to be confident that their students' test scores accurately represent (to the extent possible) student knowledge in the academic subject areas tested. According to Erwin and Wise (2002), "the challenge to motivate our students to give their best effort when there are few or no personal consequences is probably the most vexing assessment problem we face" (p. 71).

This study addresses three primary problems. The first is to what extent can student motivation be enhanced on a relatively low-interest task such as taking a low-stakes test? The second is to what extent can test performance be enhanced on a low-stakes test? Finally, what is the relationship between test-taking motivation and test performance?

Prevalence and Use of Low-Stakes Tests

The use of standardized tests to assess student academic achievement is widespread at all levels of public school education (K-12) and, to a lesser extent, in higher education. Standardized tests in K-12 are used for many purposes including assessing individual student achievement, assessing teacher effectiveness, and assessing district and state level performance (Thomas, 2005). Results from many standardized tests administered in public schools are reported to the states and U. S. Department of Education as part of the requirements for *The No Child Left Behind Act of 2001* (Geortz & Duffy, 2003; Thomas, 2005). Standardized exams are also administered in schools for a number of other reasons, the results of which never are reported to a state or federal agency. These exams are voluntarily administered by districts to monitor progress in student achievement in various subject areas, to diagnose student learning difficulties, and for other reasons.

In higher education, many universities have also adopted standardized tests to measure student success in learning academic material. The most popular tests used for this purpose include *College BASE, Collegiate Assessment of Academic Proficiency* (CAAP), *Measure of Academic Proficiency and Progress* (MAPP), and the Collegiate

Learning Assessment (CLA) (published by the University of Missouri's Assessment Resource Center, ACT, ETS, and Rand Corporation, respectively). Universities often use data from these tests to assess student learning and the effectiveness of their general education programs (Banta, 2002).

Though distinct in many ways, the use of standardized tests in elementary, secondary, and higher education share at least one common characteristic: they are typically low-stakes tests for the test takers. In other words, there is little or no meaningful consequence *to the examinee* with regard to test performance. There is concern, however, regarding the validity and usefulness of data collected under low-stakes test situations (Banta, 2002; Wainer, 1993).

Test Consequences

Understanding the consequences of a test is the critical element in deciding whether the test is a low or high-stakes test (DeCesare, 2002; Goertz & Duffy, 2003). For instance, because of the *No Child Left Behind Act of 2001*, all states must establish statewide standards for student achievement in the elementary and secondary grades (Linn, 2003). The results from the statewide assessments determine which districts/schools are making "adequate yearly progress" (Linn, 2003; Sanders, 2003). From a policy perspective, the consequences of these types of assessments are high stakes for schools and districts because district and school personnel are held accountable for the test results, often with financial implications (Goertz & Duffy, 2003). However, unless there is some meaningful consequence to the examinee (e.g., minimum score required to graduate from high school), the test is low stakes from the perspective of the

student. For example, in Missouri, students in elementary and secondary public schools take the *Missouri Assessment Program* (MAP) test (Missouri Department of Elementary and Secondary Education, 2006). The results of this test are used to evaluate the "quality" of the school district and its teachers. It is also used to characterize schools as "failing" if test scores are chronically low. The results are also reported to the federal government as part of its "Academic Yearly Progress" report. This seemingly high stakes test, however, is low stakes for the student. There is no meaningful consequence to students other than a report that they receive approximately *four months* after taking the exam. This has led to some concern that students do not take the test seriously. One Missouri parent said, "You've heard the rumors that kids taking the MAP tests don't think they're important and may not try as hard as they could" (Heavin, 2006, March 21). Conversely, a high-stakes test for students has at least some academic or other meaningful consequence to the student. For example, at present 22 states have an exit exam for high school students (Kober, Chudowsky, Chudowsky, Gayler, & McMurrer, 2006). These states require students to achieve a minimum cut score in order to graduate from high school. Clearly, these states are employing high stakes testing for their high school students: there is a clear and unambiguous consequence to students based on their test score. Another example is the SAT or ACT exam for those students wishing to go to college. These exams have meaningful consequences to students who take them: they are trying to get into college and their performance on the exam matters.

Overall in K-12 and higher education, there is concern that students experience low motivation to do well on low stakes exams because there is no meaningful consequence for the examinee to perform well (Baumert & Demmrich, 2001; Banta,

2002; Palomba & Banta, 1999; Wainer, 1993). The possibility of low motivation raises the concern of whether data collected are a valid measure of student achievement due to increased measurement error (Wainer, 1993).

All test scores have some amount of measurement error that undermines the validity of the test data (Anastasi & Urbina, 1997). Error causes individual observed test scores to be either higher or lower than the theoretical "true" score (Anastasi & Urbina, 1997). The primary consideration is to determine what level of error is acceptable. Ideally error will be minimized so that the test-takers' observed scores are as close to their true score as possible. Researchers have identified many sources of error that can lower a student's test score, low motivation being one of them (Anastasi & Urbina, 1997; Gronlund, 2006). Other factors that can introduce error in the observed score include, not allowing enough time for the test, student sickness on the day of testing, and content sampling error (Reynolds, Livingston, Willson, 2006). Researchers are now beginning to investigate the impact and relationship of low test-taking motivation and test performance as one source of error in test measurement (Brown & Walberg, 1993; DeMars, 2000; Wainer, 1993; Wolf & Smith, 1995).

CHAPTER II: LITERATURE REVIEW

Test-taking motivation

Pintrich and Schunk (2002) defined motivation as, "the process whereby goal-directed activity is instigated and sustained" (p. 5). Test-taking motivation therefore, can be defined as, "student's engagement and expenditure of energy toward the goal of attaining the highest possible score on the test" (Wise & DeMars, 2005, p. 2).

In general, research regarding test-taking motivation has taken a fairly broad view of this definition by including varied measures of motivation such as effort, interest, usefulness, value, and importance (Cole & Bergin, 2006; Wise & DeMars, 2005). Test-taking motivation research is a relatively new field of study with the majority of studies conducted in the past 10 years. Even so, one consistent finding from these studies is that, broadly speaking, motivation is correlated with test performance. For instance, O'Neil, Sugrue, and Baker (1995/1996) found that student "effort" was significantly correlated with grade 8 and 12 math performance on the National Assessment of Educational Progress (.24 and .22 respectively). Karmos and Karmos (1984) also reported that student motivational attitudes were significantly correlated with performance on the Stanford Achievement Test for grades 6 through 9. Wise and DeMars (2005) reviewed the results of 25 empirical studies investigating test-taking motivation. They found that for all but one of those studies, there was a consistent and positive difference in test scores between low and high motivated test-takers. For the 24 studies, the average effect size (*g*) was .59, leading Wise and DeMars to claim that, "motivated students perform, on average, more than one-half standard deviation higher than unmotivated students" on a low-stakes exam (p. 5).

Another fairly consistent finding is that test consequences (low and high stakes) are correlated with test-taking motivation and/or test performance. That is, students score higher on high stakes exams, compared to performance on low-stakes exams. For example, Cole and Osterlind (2005) reported that students who took the *CollegeBASE* exam as a high stakes entrance exam for admittance into a teacher education program scored significantly higher than other comparable college students who took the same test as a low stakes exam. Specifically, Cole and Osterlind found significant differences on all four subject areas tested (English, mathematics, science, and social studies), though the effect size was small for all but the English subtest (which had a moderate effect size). Wolf and Smith (1995) also investigated the effects of test consequence on test performance. In their experimental study, students were randomly assigned to one of two groups on day one of testing. On day two of testing, students switched groups so that all students participated in both groups. The consequence group was told that their performance on the test would be counted as part of their course grade. The no consequence group was told that the test would not count as part of their course grade. The consequence group reported a significantly higher level of test-taking motivation compared to the non-consequence group. The effect size was large at 1.45. The consequence group also significantly outperformed the no consequence group on test scores; however the effect size was much smaller at .26. The researchers found that the correlation between motivation and test performance was higher for students in the consequence group as compared to the no consequence group ($r = .351$ and $r = .232$, respectively). Napoli and Raymond (2004) also investigated the effect of test consequence (graded vs ungraded exam) on test performance. They reported that there

was a significant difference in test performance between the consequence and no consequence groups, with an effect size of 1.27. Other researchers who also studied the impact of consequences on test performance reported similar results (Smith & Smith, 2006; Wise, 2006; Wolf, Smith, & Birnbaum, 1995)

However, not all studies concluded that an increase in consequences led to increased scores or motivation on low stakes tests. In one of the most recent studies, Sundre and Kitsantas (2004) investigated the interdependence of self-regulatory strategies and test-taking motivation for exams that were either essay or multiple choice and that had either consequence or no consequence to the examinee. They reported that test-taking motivation was not a significant predictor of test performance for exams with consequences. Self-regulation strategies, however, were significant predictors of test performance for exams with consequences. Motivation was a significant predictor, however, for exams with no consequences. They explain this result by claiming that for consequential exams, the average score on the motivation scale is high with a low standard deviation. Essentially, most of the students display uniformly high levels of motivation (i.e., ceiling effect). However, for the non-consequential groups, motivation played an important role in predicting test performance. The overall motivation scores for the no consequence groups were lower than the motivation for the consequential groups, with much greater variability in the data. Test scores for the non-consequential groups were also lower as compared to the consequential groups. In other words, it appears that low motivation can undermine test performance for non-consequential exams; whereas high motivation for consequential testing does not seem to contribute to test scores above and beyond that explained by self-regulatory strategies.

Prior research has clearly demonstrated that students generally exhibit lower motivation for a low stakes exam and that low motivation is associated with decreased test performance. One possible explanation could be that students experience low-stakes exams as a low interest task. Interest is generally characterized as a preference for an object or task (Bergin, 1999). Interest is distinct from motivation, however, in that motivation is generally concerned with "an inner drive to action" (Bandura, 1986, p. 243), whereas interest refers, "to a person's interaction with *specific* class of tasks, objects, events, or ideas" (Krapp, Hidi, & Renninger, 1992, p. 8). Krapp et al. (1992) go on to claim that, "such specificity distinguishes individual interest from other psychological concepts such as intrinsic motivation, attention, arousal, curiosity, and exploration" (p. 8). Though distinct, motivation researchers have long claimed that interest is an important predictor of motivated behavior (Pintrich & Schunk, 2002; Wigfield & Eccles, 2000). In the context of test-taking motivation, low interest in the exam may undermine test-taking motivation and test performance.

Interest is generally characterized as either personal or situational (Bergin, 1999; Renninger & Hidi, 2002). Personal (or individual) interest refers to, "a person's relatively enduring predisposition to re-engage and persevere in work with particular content over time" (Renninger & Hidi, 2002, p. 174). Situational interest is defined as attitudes such as enjoyment that are activated in the moment or situation (Bergin, 1999; Renninger & Hidi, 2002). Both types of interest are associated with motivated behavior (Pintrich & Schunk, 2002).

In the contexts of low-stakes exams, it is proposed that students will typical experience low situational interest (interest that is activated due to the testing situation) due to the non-consequential nature of the exam.

Very few test-taking motivation studies have reported the interest level of the test-taker. However, there is some evidence that student's view low-stakes tests as a low-interest activity. For instance, results of a study recently completed by Cole and Bergin (2006) indicated that students who took a college-level general education exam as a low-stakes test reported generally low interest in the activity (M=1.97, std = 1.47, on a 0-6 point scale). Given the paucity of research in this area, however, one can only guess that student interest to do well on other low-stakes exams is low. Certainly the results reported by Cole and Bergin though, support the suspicions of many educators and educational researchers who observe students taking low stakes exams (Paris, Lawton, Turner, & Roth, 1991; Olsen & Wilson, 1991).

Thus, the focus of this study is two-fold: (a) how to effectively enhance student motivation for a low interest task such as a low-stakes test, and (b) how to increase test performance on low-stakes exams. A line of research using Self-Determination Theory has emerged directed at the challenge of enhancing motivation for low interest tasks (Ryan & Deci, 2002).

Self-Determination Theory

Self-Determination Theory (SDT) claims that three innate psychological needs motivate human behavior: relatedness, competence, and autonomy (Deci & Ryan, 2002). Relatedeness refers to one's sense of feeling connected to other individuals and/or groups

of people (e.g., a community). Competence refers to one's feeling of confidence to interact effectively in one's own environment. Autonomy concerns the source (locus) of behavior as being either internal (e.g., freely choosing to participate in a task as a result of interest) or external (e.g., required participation in a task). Cumulatively, the human desire to satisfy these needs is what energizes, sustains, and directs behavior (Deci & Ryan, 2002). According to SDT, it is through interaction with nurturing and supportive factors in our social environment that these three psychological needs (relatedness, competence, and autonomy) are fostered and enhanced.

As a sub-theory of SDT, Cognitive Evaluation Theory (CET) was proposed by Deci and Ryan (1985) to explain how the social factors that contribute to one's feeling of competence, autonomy, and relatedness produce variability in intrinsic motivation. Intrinsic motivation is defined as, "the doing of an activity for its inherent satisfactions rather than for some separable consequence. When intrinsically motivated a person is moved to act for the fun or challenge entailed rather than because of external prods, pressures, or rewards" (Ryan & Deci, 2000, p. 56). Intrinsically motivated behaviors are characterized by individuals who participate in activities not because of an external reward (e.g., money) or at the behest of others, but instead because the activity itself satisfies innate psychological need for competence, relatedness, and autonomy. In other words, the activity is the reward.

According to SDT, the need for competence and autonomy are particularly sensitive to changes in the social environment (Ryan & Deci, 2000). Social factors can either enhance or undermine feelings of competence within an individual. For example, teachers who provide a student with positive performance feedback shortly after the

completion of an academic task will undoubtedly help to enhance the students' sense of competence. Perceptions of autonomy are also particularly sensitive to one's environment. For instance, teachers who have highly regulated and structured classrooms may well undermine a student's sense of autonomy if students in such a classroom find they have little opportunity to exercise choice or make decisions regarding their learning. Therefore the social environment plays a key role in one's quest to satisfy the need for autonomy and competence. As stated by Ryan and Deci (2000a), "for a high level of intrinsic motivation people must experience satisfaction of the needs both for competence and autonomy" (p. 58). Feelings of enhanced competence or autonomy alone are not enough.

However, expecting students to be intrinsically motivated toward all academic activities, though highly desirable and ideal, may not be a realistic goal for educators and educational institutions. In educational settings, it is common to find students engaging in uninteresting tasks (Baines & Stanley, 2003). One of these activities likely includes taking low stakes tests with little or no meaningful consequence to the examinee. In situations where individuals express little or no interest in the task, it is reasonable to assume that there is no intrinsic motivation to undermine (Deci, Koestner, & Ryan, 2001). The question for educators then becomes, "How do I motivate my students to try hard on a task that is boring to most students?" SDT claims that intrinsic motivation can only be found when the activity holds some inherent interest or appeal by the individual (Ryan & Deci, 2000). The principles of CET do not apply for those activities that do not hold such inherent interest or appeal (Ryan & Deci, 2000). Instead, for such activities, we instead need to look at behaviors that are externally regulated.

Organismic Integration Theory (OIT), another sub-theory within SDT, specifically explains an individual's motivation to participate in activities that hold no inherent interest, enjoyment, or appeal to most individuals (Deci et al., 1994; Deci & Ryan, 2002; Joussemet, Koestner, Lekes, & Houlfort, 2004). In other words, these activities are said to be inherently "low interest" to most individuals. According to the theory humans tend to internalize externally regulated events when participating in an uninteresting activity (Deci & Ryan, 2002). A person participating in an uninteresting activity at the behest of others will tend to internalize the regulation of the activity to the extent that they feel a sense of relatedness of the significant other, a sense of competence in completing the task, and perceive at least some autonomy (Deci & Ryan, 2002).

Similar to CET, Organismic Integration Theory claims that this process of internalization follows a continuum of levels that are influenced by social factors (Figure 1). There are six primary levels of internalization. The lowest level is *amotivation*. This level is characterized by the *absence of any intention to act* or engage in an activity. Amotivation can result from many factors including perceiving a complete of lack of control regarding outcomes, low sense of efficacy, and/or low value for the outcome of the task (Deci, 1975; Deci & Ryan, 2000). For example, "Johnny," is a 10th grade student who chronically does not turn in his math homework. Johnny is a very low achiever in math and likely perceives no control over the successful completion of his math homework. In addition, as a result of years of negative experiences in school, Johnny has very low efficacy and value for learning math. Johnny is clearly amotivated to learn about math.

Figure 1. The Self-Determination Continuum.

Type of motivation	Amotivation	Extrinsic motivation				Intrinsic motivation
Type of regulation	Non-regulation	External regulation	Introjected regulation	Identified regulation	Integrated regulation	Intrinsic regulation
Perceived locus of causality	Impersonal	External	Somewhat External	Somewhat Internal	Internal	Internal
Quality of behavior	Nonself-determined	.....				Self-determined

Adapted from Deci and Ryan (2002) and Ryan and Deci (2000b)

The lowest level of motivated behavior is *external regulation*. This level is characterized by the lowest levels of autonomy and self-determination where individuals are motivated solely by external contingencies such as to receive a reward or avoid punishment. Externally regulated behavior leads individuals to perform (or avoid) activities based solely on the expected external contingencies.

Continuing with the example above, Johnny's 10th grade geometry teacher has given him an ultimatum: complete all the remaining math homework or be suspended from school. This form of external regulation results in Johnny completing all the remaining math homework to avoid punishment. If the threat of punishment is removed,

we could expect Johnny going back to his previous behavior of not completing his homework assignments.

The next level of internalization is *introjected regulation*. Introjected internalization is characterized by low levels of self-determination and largely external locus of causality. Individuals accept some value in the activity, but do not fully identify with the activity or believe in its utility. According to SDT, the source of introjected regulation is mostly external, with some mild internal feelings such as guilt or obligation. This level of internalized behavior in our fictitious student "Johnny" may look like this: As the school year progresses, Johnny starts to realize that his poor math performance reflects poorly on his teacher who he hears is starting to be labeled as a "bad" teacher. Though he still hates doing math homework, he disagrees with others who claim that his poor performance is the fault of his teacher. Johnny finds himself trying a little harder to do better in her math class out of a sense of guilt or obligation to his teacher.

Identified regulation is a highly-self determined form of motivated behavior, which involves the individual having higher value for the activity. At this level, individuals now perceive the personal importance and value of a task and will perceive a largely internal locus of causality. Individuals experiencing identified regulation recognize that the task is externally regulated, but they have adopted the goals, values, and importance of the task as their feelings toward the task. As a result, individuals experiencing identified regulation perceive a high amount of autonomy and self-determination.

As spring arrives, Johnny has been consistently trying harder in his geometry class. His teacher has noticed and has expressed to him additional support and

encouragement. In addition, Johnny has started to realize the importance of learning geometry. He works with his older brother building houses and he is using he new-found math skills to determine the angles of roof lines and staircases.

The final level of extrinsic motivation is *integrated regulation*. According to Ryan and Deci (2002), "integrated regulation provides the basis for the most autonomous form of extrinsically motivated behavior" (p. 18) and involves the individual perceiving the locus of causality as being entirely internal. At this level, individuals have fully adopted the values, goals, and importance of this task as being their own. As a result, integrated regulation is nearly indistinguishable from intrinsically motivated behavior.

By the end of the school year, Johnny is working hard for his brother building houses. A job and activity he really enjoys. He has also come to realize how important and useful math is for many aspects of house-building. By the end of the year, he has decided that he really enjoyed his geometry class and especially his teacher who supported and encouraged him. He will be a junior next year and his looking forward to taking the next required math class.

Intrinsically motivated behaviors are performed for the inherent interest, joy, or satisfaction of the activity. This type of motivation is entirely self-determined and autonomous, with no externally contingencies (e.g., rewards). Applied to our continuing example, Johnny finds himself working out geometry problems in his spare time and enjoys the challenge. Johnny's behavior could be characterized as intrinsically motivated. However, intrinsically motivated behavior may not always be present for students in all subject areas and academic tasks. As stated earlier, intrinsically motivated behavior can only be expected for activities and tasks that hold some level of inherent

interest or enjoyment for individuals. Can it be assumed that all academic subjects and academic activities are inherently interesting to students? Probably not. Though it may not be realistic to expect all students to display intrinsic motivation toward all academic subjects and activities, it can be expected that social factors can positively influence the behaviors of students who are being asked to participate in tasks that are less than inherently interesting.

In summary, there are four types of extrinsic motivation that represent a continuum based on the degree to which the subject feels autonomous and competent in their environment. The degree to which someone's engagement in an activity is autonomous and self-determined is an important distinction in that it is predictive of the quality of their engagement in that activity (Ryan & Connell, 1989). When students feel that their needs for self-determination (autonomy, competence, and relatedness) are being met, they will respond with more active and meaningful engagement in the activity, compared to students who feel their needs are not be met as a result of excessive extrinsic regulation through external contingencies (Boggiano, Flink, Shields, Seelbach, & Barrett, 1993).

Using Self-Determination Theory, Reeve (2001) identifies three motivational strategies that teachers can use to motivate their students. One is to use instructional strategies that facilitate intrinsic motivation by addressing the psychological needs that students have for relatedness, competence, and autonomy. For those classroom activities where there may be little inherent interest in the activity but that are deemed by school staff as important, a second strategy would be to incorporate external regulation, such as incentives, rewards, or consequences into instructional strategies. Lastly, instructional

strategies could facilitate identified or integrated regulation by enhancing student autonomy in the classroom, providing competence-enhancing feedback, and fostering student relatedness to each other and the teacher. Though externally regulated, these strategies would address the psychological needs of the students, and would also enhance their active engagement in activities that may not be inherently interesting. Though not explicitly stated by Reeve, the latter two strategies could have possible implications for enhancing test-taking motivation. Test-taking strategies to enhance motivation could include the use of rewards and/or strategies (second strategy) or strategies for facilitating internalization of the test-taking task (third strategy).

Social Context and Low Interest Tasks

To facilitate identified or integrated motivation, students must feel autonomous in their decisions as to why they are engaging in an activity. According to the SDT, there are at least three external factors that may facilitate integrated internalization, the highest form of externally regulated behavior (Deci & Ryan, 2002; Reeve, Jang, Hardre, & Omura, 2002). One is providing a personally meaningful rationale for the low interest task to the individual who is being asked to participate in the activity. The rationale helps to enhance one's value for the task. The second is acknowledging that the task is low interest. This acknowledgement helps to reduce the internal conflict the individual may have regarding his or her participation in the activity (Deci, Eghrari, Patrick, Leone, 1994). This internal conflict arises when a person is asked to do something that he or she does not want to do. Acknowledging low interest reduces the tension caused by this conflict. Third is to provide the individual with some sense of choice or control in how

he or she participates in the task. This is done by avoiding the use of controlling language such as "have to" or "you must." Individuals participating in an activity, who perceive some choice or control over the task, will tend to report higher level of internal locus of causality.

Several studies have shown the efficacy of interventions targeted at facilitating autonomy, relatedeness, and competence, thus promoting positive and effortful engagement in uninteresting activities (Jang, 2003; Deci et al., 1994; Joussemet, Koestner, Lekes, & Houlfort, 2004). For example, in an experiment conducted by Deci et al. (1994), researchers manipulated rationale, acknowledgement, and choice as independent variables. The primary dependent variable of interest was free-choice engagement time with an uninteresting computer activity. Deci et al. (1994) claimed that freely choosing to engage an activity is a measure of internalization and is similar to other studies that have used engagement time in a free-choice activity to measure intrinsic motivation (e.g., Anderson & Rodin, 1989; Ryan, Koestner, & Deci, 1991). The researchers hypothesized that if test administrators provided a rationale for engaging in the activity ("Doing this activity has shown to be useful . . ."), acknowledged that the activity was "not much fun" and "boring", and used language that conveyed choice ("If you are willing to continue. . ."), then the subjects would be more likely to engage in the activity during a free activity period. Their results indicated that collectively all three factors facilitated internalization (measured as "engagement time") significantly more than any one factor alone or the control group (no factors). For example, the average engagement time for those who experienced all three factors was 85 seconds, compared to engagement time of 55 seconds for those who experienced no facilitating factors.

Deci et al., (1994) also collected self-report data from the subjects regarding their interest, perceived usefulness, and perceived choice for the activity. These data were used to determine the extent to which the behavioral measure of engagement time indicated introjected regulation or integrated regulation. They hypothesized that subjects in the treatment group who were given choice, rationale, and acknowledgement, would report higher levels of interest, usefulness, and perceived choice that would also be positively correlated with engagement time. If so, this would indicate internal coherence with the behavioral data collected and indicate integrated regulation. Likewise, for the group that did not receive any of these facilitating social cultural factors, it was hypothesized that there would be a zero or negative correlations between these factors and engagement time. This would then indicate a lack of internal coherence and introjected regulation. Their results supported their hypotheses. For those subjects in the treatment group who were provided choice, rationale, and acknowledgement, there was a positive and significant correlation between engagement time and interest, usefulness, and perceived choice. Individuals who were not provided with a rationale, choice, or statement of usefulness had significant negative correlations with engagement time. According to the researchers, the results of these two experiments provide further evidence of the facilitating effects these social cultural factors for internalization of uninteresting activities.

More recently, Reeve, Jang, Hardre, and Omura (2002) investigated the facilitating effects of externally provided rationale on effort toward an uninteresting activity. The study included three experimental groups (external, introjected, and identified regulation group) and one control group. All subjects in the study were college

students. All the groups were presented with the uninteresting task of learning conversational Chinese. Based on their motivational mediation model, the rationale for all three interventions were variations on "reason to try." The reason to try that was presented to the introjected regulation group was that this is what you "ought" to do. The identified regulation group was presented with a rationale that explained the usefulness of the activity, while acknowledging the negative affect ("the information so far has been difficult and at times frustrating") (Reeve et al., p. 189). These reasons to try presumed to create varying degrees of external regulation for the upcoming test that would test their knowledge of conversational Chinese. Their results indicated that the identified regulation group reported significantly higher levels of importance, self-determination, and effort compared to the other two experimental groups and the control. They found no significant difference between the external, introjected, and the control group with regard to importance, self-determination, and effort.

In summary, these studies point to the positive effects of providing a rationale, acknowledgement, and choice to enhance identified internalization of uninteresting tasks. The level to which students internalize an externally regulated, low interest task will predict their level of effort directed toward the task and their subsequent performance. However, some external contingencies (e.g., tangible rewards) are used to enhance engagement in a uninteresting task, but not by facilitating a student's sense of autonomy, relatedness, or control.

Rewards

Operant Theory contends that behaviors are motivated by extrinsic rewards, not by the satisfaction of inner psychological drives. This is in stark contrast to Self-Determination Theory that claims that human behaviors are motivated by the desire to satisfy inner psychological needs (Deci & Ryan, 2002). There are many examples of how the principles of operant conditioning are used in the classroom to shape and motivate behavior (Ormrod, 2005). Some of these examples include "token economy," behavior modification, removal punishment (e.g., loss of privilege), and many others. These techniques are often effective in classroom settings. The use of rewards has also been used to enhancing test-taking motivation and performance. However, the effectiveness of using extrinsic rewards to maximize student effort during testing time to ensure the test scores are not unduly influenced by low motivation is mixed.

Two large, often-cited studies were conducted by Harold O'Neil and colleagues at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). One study involved test-taking motivation for Third International Mathematics and Science Study (TIMMS) and the other National Assessment of Educational Progress (NAEP). Both the TIMMS and NAEP are large-scale, low stakes exams.

In the NAEP study (O'Neil, Sugrue, & Baker, 1995/1995), the basic question was, "Do National Assessment of Educational Progress (NAEP) findings accurately represent student achievement, or are some students not motivated to perform well on NAEP administrations?" (p. 135). In that study 8th grade students were randomly assigned to one of three treatment groups. These groups included a financial incentive

group (reward was $1 for every correct item on exam), an ego instruction group (students are told the goal of the test is to compare students' mathematical ability with each other), and task instruction group (students are told the goal of the test is to provide opportunity for personal accomplishment). They reported that the financial incentives group scored significantly higher on the exam compared to the other three groups (ego, task, and control). There were no significant differences between the other groups. The difference between the financial incentive group and the control translates to an effect size of .41. The financial incentives group also reported significantly higher levels of effort than the other groups. However these differences were only found when students could correctly identify what group they were in. In other words, for those students who did not remember that there was an incentive (or differences in test administration instructions), there was no main effect on test score or motivation. The average financial pay out was $25 (out of a possible $41).

The study was replicated using grade 12 students. The treatment groups were the same as grade 8, with the addition of a certificate of accomplishment group (a certificate to be included in their academic transcript that is sent to college admissions offices). Unlike the grade 8 students, O'Neil et al. (1996) reported that no treatment effects of money were found on test performance or effort. They surmised that 12th grade students were not highly motivated by the reward of $1 for each correct item. They concluded that NAEP data very likely under-represent student achievement due to low motivation. The researchers go on to claim that NAEP scores, "represent what students will demonstrate with minimal effort" (O'Neil, et al., 1996, p. 135). Given the importance of NAEP test

results in evaluating the overall quality of education in the United States, these results concerned many and let O'Neil and his colleagues to further investigate.

The TIMMS study, O'Neil and his colleagues assigned 12th grade students to one of two groups (O'Neil, et al., 2005). The control group received no financial incentive. The experimental group received $10 for every correct item on the exam (compared to the $1 for each correct item in their previous NAEP study). The average student in the incentive group received $100 (average of 10 items correct). It was hypothesized that the incentive group would outperform the control group with regard to test score. Their results however, did not support this hypothesis. There were in fact no differences in test scores between the two groups. There were however, significant differences in test-taking effort. The incentive group reported much higher levels of effort. Given the results of the NAEP study, these results were construed as somewhat puzzling by the researchers. They concluded that test-taking effort was not related to test performance, but that more research is needed to understand why (O'Neil, Abedi, Miyoshi, & Mastergeorge, 2005). They also concluded that large scale low-stakes assessment can move forward with more confidence because incentives seem to have little or no effect on test scores.

Aside from assessment situations, in other areas of education, the use of rewards to entice students to engage in an activity is common place in educational settings (Good & Brophy, 2003; Pintrich & Shunk, 2002). This is especially true when teachers are trying to motivate students to participate in an activity that is low interest.

However, according to Self-Determination Theory, the use of external instrumentalities (e.g., performance contingent reward) undermines the development of

student's autonomy, regardless of how interesting the task may be (Deci & Ryan, 1985). In addition, the use of external rewards fosters a dependence on the reward for future task engagement. In general, SDT researchers claim that the use of rewards to gain compliance or facilitate effortful engagement in a boring task is not supported by past research. In 2001, Deci, Koestner, and Ryan published a meta-analysis of 13 studies that investigated the effects of tangible rewards (e.g., prizes, gold stars, student awards, etc) on interesting and boring tasks. They investigated the effects of these tangible rewards on free-choice measures of intrinsic motivation to participate in an uninteresting or dull activity. Not surprisingly, they found that tangible rewards "do not undermine people's intrinsic motivation for dull tasks because there is little or no intrinsic motivation to be undermined" (Deci, et al. p. 14). However, they also found that rewards did not enhance or facilitate motivation to engage in an uninteresting activity.

In a follow-up study since the publication of the meta-analysis by Deci, Koestner, and Ryan (2001), Joussemet et al. (2004) investigated the use of tangible, engagement-contingent reward and autonomy support to enhance a child's self-regulation for an uninteresting task. Children in the reward groups all received a yo-yo. The autonomy supportive groups were provided instructions that conveyed choice ("If you choose to continue. . ."). In addition, these groups were provided a rationale (usefulness of the activity) and feelings of negative affect were acknowledged ("You might find this task a little boring. . ."). To convey external control, instructions to the children in the non-autonomy supportive groups included words such as, "have too," "should," etc. A total of four groups were created: autonomy supportive instructions, autonomy supportive instructions with reward, control instructions, and control instructions with reward. The

outcome variables for this study were free-choice persistence in the activity, self-reported affect, and subject's perception of the activity's task value as it relates to the internalization of the uninteresting activity. They found the autonomy support group (with and without the reward) reported significantly higher levels of positive affect for the dull activity. However, they found no treatment main effect with free-choice engagement and task value. The researchers attribute this to the possibility that the task was too dull and no amount of intervention was going to dramatically affect engagement time or affect.

However, the researchers did find that those children who received no reward indicated higher levels of internalization and integration compared to children offered a reward. Children in the no reward group had a higher correlation between free-choice engagement time and affect ($r = .38$), compared to children who received a reward ($r = -.33$). In other words, as predicted by SDT, autonomy support predicted integrated functioning whereas rewards promoted alienated functioning. As stated by Joussemet et al (2004), "There is evidence that autonomy support promotes active engagement with uninteresting but important activities" (p. 142). Overall Joussemet et al. (2004) concluded that the effects of rewards were negligible for uninteresting activities. There was some evidence, however, that autonomy support facilitated the child's ability to integrate an uninteresting task.

However, not all rewards are the same. According to Deci (1975), rewards can be characterized as either controlling or informational. According to SDT, people often perceive extrinsic rewards (e.g., tangible rewards such as money) as a controlling strategy that contributes to a perception of an external locus of causality (Ryan & Deci, 2000).

This perception of an external versus internal locus of causality results in decreased autonomy and decreased perceptions of personal control of behavior. As a result, "Students who are overly controlled not only lose initiative but also learn less well, especially when learning is complex or requires conceptual, creative processing" (Ryan & Deci, 2000, p. 59). Controlling rewards are often perceived as a "means to an end." Once the reward for the activity is removed, the individual's motivation to participate in that activity is diminished (i.e., overjustification effect). Informational rewards however, provide individuals with feedback regarding their performance in a way that is non-controlling and facilitates their feelings of competence and autonomy (Deci et al., 1994). Controlling and informational rewards can be either verbal or tangible (Good & Brophy, 2003). For instance, intrinsic motivation or internalization of externally regulated activities is likely to be enhanced if verbal feedback is given regarding student performance that is positive and non-controlling. Likewise, verbal feedback that is presented in a controlling manner and conveys a message of performance judgment will likely decrease intrinsic motivation. According to SDT researchers, tangible rewards are almost always perceived as controlling, though it is theoretically possible for tangible rewards to be used to convey positive, non-controlling performance feedback. In summary, SDT researchers claim that there is no compelling reason to use any rewards other than those that are provided in a non-controlling, informational manner.

Thus, according to SDT, providing a rationale, choice, and acknowledgement is the most effective way to achieve integrated regulation and enhance motivation for a low-interest task. In addition, the use of informational, non-controlling performance feedback

(e.g., positive verbal feedback) would also enhance integrated regulation and motivation by promoting an individuals value for the task.

Other motivation researchers have a somewhat different perspective on the use of rewards (Good & Brophy, 2003; Pintrich & Schunk, 2002), though they tend to share the common belief with SDT researchers that the most appropriate rewards are those that are non-controlling. For instance, Pintrich and Schunk (2002) generally support the use of rewards as long as they are linked with student performance. Rewards that are contingent on student performance convey a positive message to the student regarding their learning. According to Pintrich and Schunk (2002), performance contingent rewards such as verbal praise, extra credit, and competence feedback all act to enhance students' self-efficacy and motivation to learn. As noted by Good & Brophy (2003) rewards may be appropriate especially given "motivational problems facing teachers primarily involve getting students to put forth consistent learning efforts whether they find the learning tasks interesting or enjoyable, not just maintain intrinsic motivation to engage in interesting tasks" (p. 225).

There have been no studies however, testing the effects of autonomy-supportive strategies of rationale, choice, and acknowledgement on test-taking motivation and test performance. As indicated above however, there have been studies testing the efficacy of providing both task and performance contingent rewards to enhance test-taking motivation and performance.

Conclusion

SDT's Organismic Integration Theory provides us with an understanding of why some students may be more motivated than others for a low-interest low-stakes exam. Specifically, the theory claims that if the student is presented with a rationale for taking the test, is asked to engage in the exam in a non-controlling manner, and it is acknowledged to the student that the exam is not very interesting but nonetheless important, then the student is more likely to report a sense of autonomy, competence, and relatedness toward that activity. This enhanced internalization will result in higher degree of effort directed toward the activity and enhanced performance. Also, based on previous studies by Deci et al (1994) and Joussemet et al. (2004), individuals who perceive greater levels of autonomy support for the low interest activity will report a higher positive correlation between affect toward the activity and free-choice measures of task engagement. Those individuals who do not perceive an autonomy supportive environment will experience greater degree of alienation evidenced by negative correlation between affect and free-choice task engagement.

According to the self-determination theory, students provided with a choice, rationale, and acknowledgement should report higher levels of motivation and test performance compared to other students not provided with choice, rationale, and acknowledgement. However, there is some evidence that financial, performance-contingent incentives may also facilitate students' test-taking motivation and performance. This study will investigate the differences in test-taking motivation and test score on a math achievement test due to group membership in the self-determination group, the lottery group, or a control.

Hypotheses

The specific hypotheses that will be tested in this study include:

Hypothesis 1: The autonomy supported group will report higher levels of interest, value, perceived choice (three measures indicating internalization), internal locus of causality (autonomy), test-taking effort, and total number of items attempted compared to the other two groups.

Hypothesis 2: The two experimental groups will have significantly higher scores than the control group on the math achievement test.

Hypothesis 3: There will be no difference between the three groups regarding the impact of value/usefulness, choice, interest, and effort on test performance.

CHAPTER III: METHOD

Pilot Study

Method and Findings

A pilot study was conducted to accomplish four methodological goals. First the pilot study was used to refine the math test in terms of its length and difficulty. Second goal was to refine the scripts read to each of the three groups. A third goal was to collect data regarding how interesting participants found the activity. Finally, a fourth goal was to determine the appropriateness of the lottery system proposed for this study.

Participants

Participants were 49 undergraduate students (21 female, 28 male) recruited from an introductory freshman-level psychology class at a large Midwestern public university. All participants received course credit for their participation. The study passed human subjects review.

Procedure and Experimental Conditions

The 49 students who signed up for the study were randomly assigned to one of 3 experimental conditions: control, autonomy supported, and lottery. On the date of the study, students who showed up were asked to report to one of three rooms that corresponded to their assigned group.

Once in their respective rooms, students were greeted by the test proctor. In all there were three proctors covering the three rooms. These proctors were all male, Caucasian and between the ages of 40 and 50. The study followed approved human

subjects research protocol, including gaining each student's consent to participate (Appendix A), followed by the exam and questionnaire (Appendix B). The initial instructions read by each proctor for each group were the same:

> Let's get ready to start. In front of you are three things. First thing I want you to look at is the letter. It is an informed consent indicating that you agree to participate in this study. Please read and sign. When you are done, turn it over and move it to your right so that I will know when everyone has completed this step. <<<Wait>>>
>
> Now look to your left. There should be a packet titled, "Mathematics achievement test." Please turn this over, but do not turn the page.

After these initial instructions, instructions for each group varied to reflect the treatment in the experiment. The treatment for each experimental group is described below.

1. Autonomy Supported Group: For this group, the purpose of the instructions were to facilitate the test-takers' autonomy and internalization for the task, thereby enhancing their test-taking motivation. The instructions were modified to focus students' attention on the three autonomy-supportive factors that facilitate an integrated level of internalization: meaningful rationale, acknowledgment, and choice or control over the task (Deci et al., 1994). The script for this treatment group was:

> The test you are being asked to complete is a math achievement test. As I am sure you are aware, math is very important and useful in almost any career path you may choose. A strong understanding of math will prove useful to you whether you are interested in being a teacher, biologist, social worker, engineer, nurse, or just about any type of career. This test will help you to better understand your math knowledge. I know that some people don't think math is very fun, and other people think it is kind of stressful. So I can perfectly understand that you might not be enthusiastic about this -- or maybe you are. Either way, we sincerely request that you try hard, to help us understand math skill development. You have 15 minutes to complete as much as you can. You may write on the exam to work out problems. At the end of the 15 minutes, you will be asked to stop working on the exam.

Please turn over the first page and start now.

To reinforce the instructions, summary text was included at the top of the first page of math achievement test. The top of the first page for the autonomy-supported group read:

Mathematics Achievement Test

Math is very important and useful in any career.
You may not be enthusiastic, but please try hard.

2. Lottery Group: Past research has shown the beneficial effects on performance of performance contingent rewards, including those that are lottery based (Martens, Ardoin, Hilt, Lannie, Panahon, & Wolfe, 2002; Porter, Whitcomb, 2004). Students in the Lottery Group were instructed that if they passed a cut score on the exam, they would be entered into a raffle drawing for a $100 Wal-Mart gift card. Entry into the raffle was based on a two-level cut-score that was intended to motivate all of the students, regardless of prior achievement level. The script for this group was as follows:

> The test you will be taking is a math achievement test. You have 15 minutes to complete as much of the exam as you can. Based on the score you receive on the exam, you may be entered into a lottery for a $100 gift card to Wal-Mart. To be entered into the drawing once, you must score at least 50% correct on the exam. If you score 80% or higher, you will be entered twice into the drawing. The winner will be notified by the email address you provided within 24 hours. Indicate your answers on the exam. You may write on the exam to work out problems. At the end of the 15 minutes, you will be asked to stop working on the exam.
>
> Please turn over the first page and start now.

To reinforce the instructions, summary text was included at the top of the first page of math achievement test. The top of the first page for the lottery group read:

Mathematics Achievement Test

Lottery for $100 gift card to Wal-Mart.

3. Control Group: Students in the control group were read the standard test administration script written by the test publisher. The script for this group was:

> The test you will be taking is a math achievement test. You have 15 minutes to complete as much as you can. Indicate your answers on the exam. You may write on the exam to work out problems. At the end of the 15 minutes, you will be asked to stop working on the exam.
>
> Please turn over the first page and start now.

For the control group there were no additional instructions on the top of page 1 of the exam.

After hearing the instructions, students were asked to complete a short 15-item math test and questionnaire. The math test was an abbreviated version of the mathematics portion of the *Missouri Mastery and Achievement Test* (MMAT). Students had fifteen minutes to complete this portion of the test. After students completed the test, they were asked to complete the *Activity Perception Questionnaire* (Deci, et al., 1994), as well as the Self-Determination scale, measure of test-taking effort, and demographic information. Total time for all activities was scheduled at ½ hour.

Findings

The Pilot Study had four methodological goals. The first goal was to determine if the length and difficulty of the test was appropriate. The abbreviated version of the

MMAT used in this study is a high school level math achievement test. It was assumed that a high school math test would still prove to be challenging given that students in the study were mostly freshman and that for many students math is a difficult subject. However, many students in the pilot study were able to easily finish the test in the allotted time. This indicated that the test was not long enough or difficult enough to discern low and high math achievement. It was decided for the main study to use the full MMAT math test (44 items) and allot ½ hour. The full math test is more rigorous and covers more math subjects (algebra, geometry, etc). It was felt that few if any students in the main study would be able to complete the test in the allotted time. This would provide greater differentiation in scores.

The second goal was to refine the scripts used in the study. This was accomplished by getting feedback from the three proctors (myself and two others). Prior to the session each proctor was asked to observe and take notes regarding student reaction to the script and testing procedures in general. Based on this information, some changes were made to the scripts, most notably the lottery group script and the autonomy supported group script. The primary change to the lottery group script clarified some students' confusion regarding eligibility for the reward. The change to the autonomy support group was intended to emphasize even more the choice and acknowledgement elements.

The assumption for this study is that taking a low-stakes math test is a low interest task. Therefore, the third goal was to determine whether or not this task truly was of low interest to students. Data from the control group were used to determine the interest level

in the task. The pilot study did confirm this task was moderately low interest. The mean interest for this task was 3.27 (std = 1.53) on a scale of 1 (no interest) to 7 (high interest).

Finally, the fourth goal was to determine the appropriateness of the proposed lottery system. Only 2 out of the 16 students did not make the minimum cut score of 50%. This meant that 14 of the 16 students in the lottery group were eligible for entry into the lottery for the $100 gift card. Five students received a score of 80% of better. These students were entered twice into the lottery. In addition, according to the proctor in the room, students seemed to generally understand the rules. The primary suggestion from the proctor was to point out that because the group was small; students had a relatively high probability of winning.

Summary

The findings from the pilot study and the implications for the main study are summarized as:

1) Use full version of math test; pilot study test too short and not rigorous enough
2) Refine scripts for the lottery and autonomy group. The lottery group script was refined to better explain the lottery system and eligibility for inclusion. The Autonomy script was modified to further focus participation attention on the acknowledgment and choice elements.
3) Pilot study confirmed that the task is in fact a "low interest" activity. Main study can proceed assuming task is of low interest to test-takers.

4) Finally, it was decided that the odds of winning the $100 for students in the Lottery Group should be emphasized.

Main Study

Power Analysis

Based on a power analysis using conservative parameter estimates, it was determined that at least 80 students would be needed in each group to reduce the risk of a Type II error. Therefore, for this study I recruited 240 freshman and sophomore students from the University of Missouri-Columbia.

Recruitment Procedures

An email list of all freshman and sophomore students was obtained from the university. In compliance with university procedures, students were emailed requesting their participation in the research study. Because it is difficult to recruit students, students were instructed that they would receive a $15 gift card to Wal-Mart in appreciation for this participation in an hour session. They were instructed that this gift card would be given to them prior to leaving the testing session. Each student who agreed to participate could choose one of two testing dates. Students who had signed up for that session were emailed a reminder approximately four hours before each testing time. Financial support for this study came from a grant from the University of Missouri-Columbia's Research Council. The money from this grant paid the $15 each student

received as payment for participation, paid the lottery prizes, and purchased the email recruitment list.

Participants

Over the course of two testing days, a total of 246 freshman and sophomore students participated in the study. These students comprised 187 freshman and 59 sophomores. They self-identified as 77% Caucasian, 11% Black, 6% mixed or other, and 3% Asian. In addition, these 246 students were enrolled in over 75 different majors across campus. Just over 38% of the participants were male. The average age was just under 20 years old (See Appendix C for demographic characteristics).

Procedure

Random assignment occurred on the day of testing. As students checked-in on the day of testing, they were asked to select a piece of paper from a box. The paper instructed students to proceed to one of three rooms. Unbeknownst to the student at the time, each room represented either one of two treatment groups or the control group. Students were also given a registration form to complete while they waited for further instructions. The registration form asked students to provide information needed to allow disbursement of the $15 gift card. When all students were checked in the proctor for each group read the instructions which included informed consent procedure (Appendix A), followed by the exam and questionnaire (Appendix B). The script was as follows:

> Let's get ready to start. If you have not yet completed the registration form, please put that to the side and complete it at the end of the session. In front of you are three things. The first thing I want you to look at is the letter. It is an

> informed consent letter indicating that you agree to participate in this study. Please read and sign. When you are done, turn it over and move it out of the way so that I will know when everyone has completed this step. <<<Wait>>>
>
> Now look to your left. There should be a scantron sheet. Please fill in your student ID number on this sheet. Also, where it says "Grade", please mark 13 if you are a freshman, 14 if you are a sophomore, or 15 if you are a junior. When done turn this form over. There should also be a packet titled, "Mathematics achievement test." Please turn this over, but do not turn the first page.

The next instructions varied depending on the group assignment.

1. Autonomy-Supported Group: For this group, the instructions were intended to promote the test-takers' autonomy and internalization for the task in order to enhance their test-taking motivation. The instructions were modified to focus students' attention on the three autonomy-supportive factors that facilitate an integrated level of internalization: meaningful rationale, acknowledgment, choice or control over the task. It was expected that the modified instructions would result in students reporting higher levels of autonomy. The script for this treatment group follows:

> The test you are being asked to complete is a math achievement test. As I am sure you are aware, math is very important and useful in almost any career path you may choose. A strong understanding of math will prove useful to you whether you are interested in being a teacher, biologist, social worker, engineer, nurse, or just about any type of career. This test will help you to better understand your math knowledge. I know that some people don't think math is very fun, and other people think it is kind of stressful. So I can perfectly understand that you might not be enthusiastic about this -- or maybe you are. Either way, we sincerely request that you try as hard as you can to do well on this exam so that we may better understand math skill development.
>
> You have 30 minutes to complete as much as you can. Please do not write on the exam. You may use the loose paper accompanying the exam to work out problems. Please indicate all your responses on the scantron sheet provided. At the end of the 30 minutes, you will be asked to stop working on the exam.
>
> Please turn over the first page and start now.

To reinforce the instructions, summary text was included at the top of the first page of math achievement test. The top of the first page for the autonomy-supported group read:

Mathematics Achievement Test

Math is very important and useful in any career.
You may not be enthusiastic, but please try hard.

2. Lottery Group: Past research has shown the beneficial impacts of performance contingent rewards, including those that are lottery based (Martens, Ardoin, Hilt, Lannie, Panahon, & Wolfe, 2002; Porter, Whitcomb, 2004). Students in the Lottery Group were instructed that if they passed a cut score on the exam, they would be entered into a raffle drawing for a $100 gift card to Wal-Mart. Entry into the raffle was based on the same two-level cut-score employed during the pilot study. Thus the lottery system used in the main study included a 50% level that could be achieved by most test-takers and a higher level (80%) that could only by achieved by test-takers with higher level math skills. The 50% level provided one chance in the lottery, and the 80% level provided two chances. The script for this treatment group was as follows:

> The test you will be taking is a math achievement test. You have 30 minutes to complete as much of the exam as you can. Based on the score you receive on the exam, you may be entered into a lottery for a $100 gift card to Wal-Mart. To be entered into the drawing once, you must score at least 50% correct on the exam. If you score 80% or higher, you will be entered twice into the drawing. To give you an idea of the odds, one of you in this room will win the $100. The winner will be notified by email within 72 hours.
>
> You have 30 minutes to complete as much as you can. Please do not write on the exam. You may use the loose paper accompanying the exam to work out problems. Please indicate all your responses on the scantron sheet provided. At the end of the 30 minutes, you will be asked to stop working on the exam.

Please turn over the first page and start now.

To reinforce the instructions, summary text was included at the top of the first page of math achievement test. The top of the first page for the autonomy-supported group read:

Mathematics Achievement Test

Lottery for $100 gift card to Wal-Mart.

3. Control Group: Students in the control group were read a basic test administration script that was not intended to facilitate autonomy support or elicit any feelings of external control. The script for this group was:

> The test you will be taking is a math achievement test. You have 30 minutes to complete as much as you can. Please do not write on the exam. You may use the loose paper accompanying the exam to work out problems. Please indicate all your responses on the scantron sheet provided. At the end of the 30 minutes, you will be asked to stop working on the exam.
>
> Please turn over the first page and start now.

For the control group there were no additional instructions on the top of page 1 of the exam.

After the instructions, students were asked to complete the mathematics portion of the *Missouri Mastery and Achievement Test* (MMAT) (1997). Thirty minutes were allotted for students to take this test. At the end of thirty minutes, the proctor stated:

> Please stop working on the exam. Turn over the exam and place it to your left. Now please turn over the packet to your right. It looks like this. <<<Hold up copy of questionnaire>>> Please write your student number in the space provided at the top. After you have done this, please complete this questionnaire. When done, turn over the questionnaire so that I know when everyone has completed it. If you did not get a chance to complete the registration form, you can finish it now. I will now be coming by to give you your $15 gift card. Once

you have received your card and signed the registration form, you may leave. Thank you.

Study Variables

Math Test Score

The *Missouri Mastery and Achievement Test* (MMAT) was the Missouri statewide test until 1998 when it was replaced with the *Missouri Assessment Program*. The MMAT tests students' knowledge in 4 general content areas (English/Language Arts, Mathematics, Science, Social Studies/Civics). For this study, only the 44-item Grade 10 mathematics portion of the MMAT was used. The math portion of the exam was chosen primarily because many students find math to be a difficult subject. It was assumed that many freshman and sophomore college students will still difficulty even with high school level mathematics. As a result, it is not likely that there will be a "ceiling effect" with regard to the math scores. It is likely that there will be a good deal of variability in the data. The math portion of the MMAT includes three clusters: computation and estimation; geometry and measurement; and interpretation and problem solving (Missouri Department of Elementary and Secondary Education, 1997). Percent correct will be examined to determine differences between the three groups with regard to their test performance.

Test-Taking Effort

Effort was measured using a one item measure from the Test-Taking Motivation Questionnaire (Cole & Bergin, 2006). In a study by Cole and Bergin, subject-specific effort to do well on a test was measured with one item per subject. It was found that this

one item measure was predictive of test scores, even more so sometimes than prior academic achievement (ACT or GPA). The effort item included in the study was, "On a scale of 0 to 100% where 100% means you put forth maximum effort, approximately how hard did you try on this test?" The response categories ranged from 0% to 100% in 10% increments for an 11 point scale (Appendix B).

Interest, Value, and Perceived Choice

The student questionnaire includes items from the *Activity Perception Questionnaire* (APQ), which is a modified version of the *Intrinsic Motivation Inventory* (IMI) (Ryan, 1982). The IMI was developed to assess individuals' experiences with experimental tasks. Over the years, this inventory has been successfully used in many research studies investigating intrinsic motivation (McAuley, Duncan, & Tammen, 1989; Tsigilis, & Theodosiou, 2003). The inventory includes 7 scales (interest/enjoyment, perceived competence, effort/importance, pressure/tension, perceived choice, value/usefulness, and relatedness). The instrument has been found to have adequate psychometric properties (McAuley, Duncan, & Tammen, 1989). The APQ (used in this study) is a modified version of the IMI and was developed for internalization studies that involve external regulation (Deci et al., 1994). The APQ uses three of the seven IMI scales (interest, value, choice).

Response categories for these and all scales ranged from a 1 ("Not at all true") to a 7 ("Very true"). The interest scale had a total of 8 items. An example item from this scale is, "This activity was fun to do." The value scale had a total of 9 items. An example item from this scale is, "I think this is an important activity." The perceived

choice scale had a total of 8 items. An example item from this scale is, "I did this activity because I wanted to" (Appendix B). Following the procedures of Deci et al., (1994), mean scores were calculated for each sub-scale.

I used the factor and reliability analysis in SPSS 14.0 to investigate the factorial structure and internal consistency of the interest, value, and perceived choice scales. Principal axis factoring with direct oblimin rotation was used with four factors identified for extraction. Tests of sampling adequacy (KMO = .892) and sphericity (Bartlett's < .001) all proved adequate for continuation of the factor analysis (Tabachnick & Fidell, 2000). The pattern matrix correctly identified the three unique factors. All items correctly loaded onto their intended factors (Appendix D). One item in the value scale (Q28, "I am willing to do this activity again because I think it is somewhat useful.") however did not load on any scale. This item was omitted from further analysis. The internal consistency for these three measures was adequate. The interest scale had the highest internal consistency (alpha=.942), followed by value (alpha = .859) and choice (alpha = .781).

Autonomy

The measure of autonomy was adapted from an instrument used in a study conducted by Ryan and Connell (1989). In that study, Ryan and Connell investigated children's perceived locus of causality and self-reported level of external regulation. The four levels of external regulation are external, introjected, identified, and intrinsic. An example item with the stem 'Why do you do your homework' was used with four levels in the Ryan and Connell (1989) study: "Because I will get in trouble if I don't" (external),

"Because I want the teacher to think I am a good student" (introjection), "Because I want to understand the subject" (identification), and "Because it is fun" (intrinsic) (stem: Ryan and Connell report internal consistencies for each reason category between .62 and .82, depending on the sample. To ascertain whether the children were able to discriminate between the four reasons for trying, bivariate correlations by group for each category were calculated. Reasons for trying adjacent to each other (e.g., intrinsic and identified) should have the highest correlations compared to reasons that are not adjacent (e.g., intrinsic and external). Data reported by Ryan and Connell (1989) support this pattern, that children were able to adequately distinguish reasons for trying.

In the current study, the "reasons for trying" were the same as in the Ryan and Connell study. However, specific items needed to be adapted to college students and the experimental task. For instance, for the external reason, Ryan and Connell included an item "Because I will get I trouble if I don't." In the current study, this item was modified to read, "Because that is what the experimenter says I should do." Three items were included for each reason (Appendix B).

I used the factor and reliability analysis in the SPSS 14.0 program to investigate the factorial structure and internal consistency of the four levels of external regulation scales. Principal axis factoring with direct oblimin rotation was used. Tests of sampling adequacy (KMO = .773) and sphericity (Bartlett's < .001) all proved adequate for continuation of the factor analysis (Tabachnick & Fidell, 2000). Initially, each factor contained three items. A review of the pattern matrix revealed that one item intended for the extrinsic factor also cross-loaded on to the introjected factor; however, the item did load higher on the intended factor rather than the introjected factor. It was decided to

leave this item with its intended factor. One item from the extrinsic factor loaded very low and one item from the identified factor did not load at all. In addition, the internal consistency analysis revealed that the intrinsic factor was substantially improved with the removal of one of the items. The result was four factors: one with three items (introjected) and the other three with two items (intrinsic, extrinsic, and identified) (Appendix D). The internal consistency for the intrinsic, identified, introjected, and external are .843, .754, .798, and .732, respectively.

In addition to the factor analysis and internal consistency analysis, it was also important to determine whether subjects in the study could adequately distinguish between the four reasons to try. Similar to the Ryan and Connell study (1989), students in this study were able to adequately distinguish between the reasons for trying hard during the test. Based on the bivariate correlations for each group between the four reasons, it was determined that for each group, the highest correlations were between adjacent reasons (e.g., intrinsic and identified) and lower correlations between non-adjacent reasons (e.g., intrinsic and external). A table of intercorrelations is provided in Appendix E for all variables.

Gender and ACT Test Scores

Gender was included in this study. Many previous studies have reported the observed differences in math achievement scores between males and females (Halpern, 2000). ACT test scores as a predictor of future math achievement was also included (Linn, 1990). The mean ACT score for this sample was 25.9 (std = 3.8)

Analysis of data

Hypothesis 1: The autonomy supported group will report higher levels of interest, value, perceived choice (three measures indicating internalization), internal locus of causality (autonomy), test-taking effort, and total number of items attempted compared to the other two groups.

Multivariate analysis of variance (MANOVA) will be used to test hypothesis 1. One factor will be included as an independent variable: group assignment. The dependent variables will include interest, value, locus of causality, test-taking effort, and total number of test items attempted. Main effects will be examined as well as post-hoc tests to determine differences between three groups for each of the dependent variables.

Hypothesis 2: The two experimental groups will have significantly higher scores than the control group on the math achievement test.

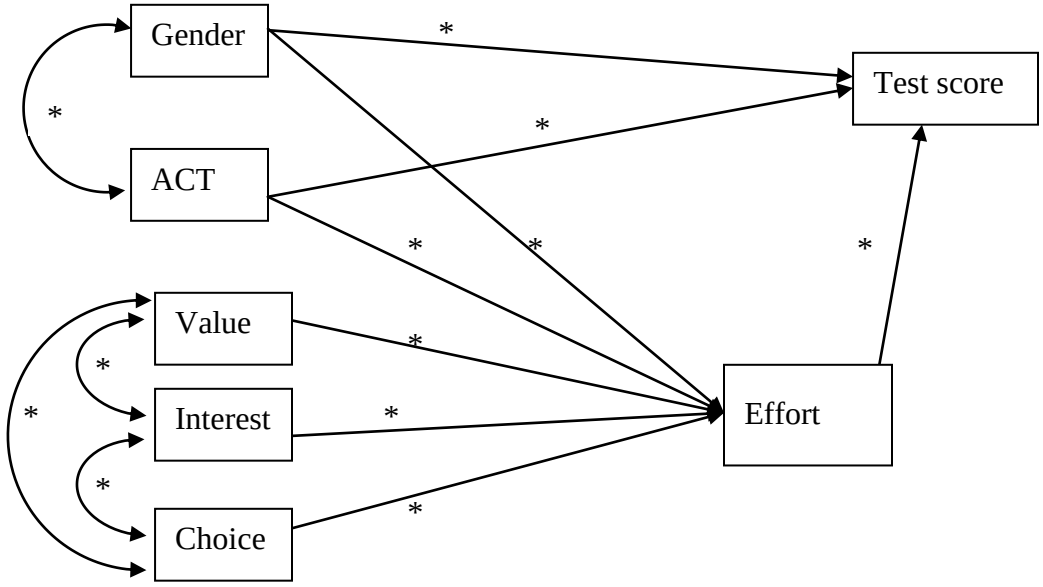
Analysis of covariance variance (ANCOVA) was used to test hypothesis 2. Two factors were included as independent variables: 1) group assignment to the reward, autonomy, or control group; and 2) gender. ACT scores were used as a covariate to control for difference in prior math achievement. The use of covariates in experimental design generally has two purposes (Tabachnick & Fidell, 2000; Bryman & Cramer, 2005). The purposes for including ACT as a covariate is to reduce the error term thereby increasing the sensitivity of the test of main effects and interactions. Another purpose is to adjust the mean scores on the math test by taking into account past academic achievement.

The dependent variable will include math test score. Main effect will be examined as well as post hoc tests to determine differences between three groups for each

of the dependent variables. In addition, test of interaction using gender will examine if males or females respond differently to the treatments.

Hypothesis 3: There will be no difference between the three groups regarding the impact of value/usefulness, choice, interest, and effort on test performance (Figure 2). Using EQS 6.1, multi-group path analysis will be used to test model invariance. The independent variables include measures of internalization (value, choice, and interest). Dependent variables include test-taking effort and test performance. Control variables of ACT and gender are also included to rule out their influence on the test scores and effort (Cole & Bergin, 2006).

Figure 2. Initial path model



CHAPTER IV: RESULTS

Preliminary Analyses

Preliminary analyses were conducted to assure that the participants did in fact find the test-taking activity as uninteresting. To confirm this, data from the control group were examined because their level of interest was not influenced by an intervention. There were a total of 80 participants in the control group. These students rated the test-taking activity as mildly uninteresting (M = 3.27, SD = 1.39, on a 7 point scale where 7 is high interest).

Effect of Rationale, Choice, and Usefulness and Reward on Internalization, Autonomy, and Test-Taking Effort

Hypothesis 1 claimed that the autonomy group should experience greater levels of interest, value, perceived choice (three measures indicating internalization), internal locus of causality (autonomy), test-taking effort, and total number of items attempted compared to the other two groups. The descriptive statistics for these variables are reported in Table 1 by group. The main effect was significant (Wilks Lambda $F=1.950$, $p<.05$). Tests of Between-Subjects revealed that there were no significant differences between groups for interest, value, choice (measures of internalization) or intrinsic, identified, or introjected regulation (autonomy). Significant differences were found between the groups with regard to number of items attempted ($F=4.371$; $p<.05$), test-taking effort ($F=5.833$; $p<.01$), and external regulation ($F=7.203$, $p<.001$). No gender interactions were found.

Table 1. Means and Standard Deviations

	Group	n	Mean	Std. Deviation
Interest	Autonomy	80	3.31	1.28
	Control	73	3.27	1.39
	Lottery	77	3.42	1.52
Value	Autonomy	81	3.82	1.16
	Control	78	3.83	1.44
	Lottery	81	3.88	1.27
Choice	Autonomy	78	6.27	0.71
	Control	77	6.17	0.85
	Lottery	82	6.22	0.87
Intrinsic regulation	Autonomy	82	2.72	1.58
	Control	80	2.79	1.68
	Lottery	84	2.73	1.57
Identified regulation	Autonomy	82	4.51	1.72
	Control	80	4.47	1.63
	Lottery	84	4.50	1.68
Introjected regulation	Autonomy	82	5.09	1.45
	Control	80	4.96	1.53
	Lottery	84	5.33	1.32
External regulation	Autonomy	82	4.97	1.56
	Control	80	3.84	1.66
	Lottery	84	4.36	1.79
Test-taking effort	Autonomy	81	8.67	1.78
	Control	80	7.95	1.50
	Lottery	82	8.68	1.52
Total number of test items	Autonomy	82	37.82	6.21
attempted (out of 44)	Control	80	37.43	6.11
	Lottery	84	39.80	5.57

Scheffe's post-hoc tests revealed that there was a significant difference between the lottery group and the control with regard to number of items attempted, effort, and perceived external regulation (Table 2). The lottery group attempted an average of 2.7 more items on the exam than the control. In addition, the lottery and autonomy groups reported significantly more effort toward the exam compared to the control group. The autonomy group also reported significantly higher levels of external regulation compared to the other two groups. This result is somewhat perplexing and counter-intuitive and is further discussed below. These results only partially support hypothesis 2. The autonomy group did report significantly higher effort compared to the control group, but they also reported significantly higher external regulation compared to the other two groups.

Table 2. Scheffes' Post-Hoc Test Results

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)	*p*.
# of items attempted	Autonomy	Control	.40	.919
	Autonomy	Lottery	-2.30	.065
	Control	Lottery	-2.70*	.025
Effort	Autonomy	Control	.604*	.032
	Autonomy	Lottery	-.138	.834
	Control	Lottery	-.742**	.007
External Regulation	Autonomy	Control	1.0381***	.001
	Autonomy	Lottery	.4341	.291
	Control	Lottery	-.6040	.097

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

Effect of Rationale, Choice, and Usefulness and Reward on the Math Achievement Test

Hypothesis 2 claimed that the two experimental groups would have significantly higher scores than the control group on the math achievement test. Test performance was calculated as the number of items correct divided by the total number of items in the test. A 2x3 factorial design was used to investigate differences. Two factors were included as independent variables: 1) group assignment to the reward, autonomy, or control group; and 2) gender. ACT scores were included as a covariate to control for prior math achievement.

Main effects were found for group ($F=5.638$; $p<.05$), gender ($F=7.209$; $p<.01$), and interaction ($F=2.968$; $p<.05$). Pairwise results for group, gender, and interaction are reported in Tables 3, 4, and 5. As shown in Table 3, the lottery group scored significantly higher on the exam compared to the autonomy group ($p<.01$). There were no other significant differences in test scores between the groups. These results partially support Hypothesis 2. When controlling for ACT scores, the lottery group scored significantly higher on the test compared to the other two groups, whereas there were no significant differences in test scores between the autonomy and control. However, this main effect should be interpreted cautiously because there was an interaction, which is discussed next.

Table 3. Adjusted Mean Test Performance Scores by Group

Group	n	Mean	Std. deviation
Autonomy	78	67.60	1.36
Control	75	71.13	1.34
Lottery	77	73.85	1.28

As indicated in Table 4, males across all 3 groups scored an average of about 4 points better on the exam compared to females ($p<.01$). This result is consistent with other research examining gender differences in math performance.

Table 4. Adjusted Mean Test Performance Scores by Gender

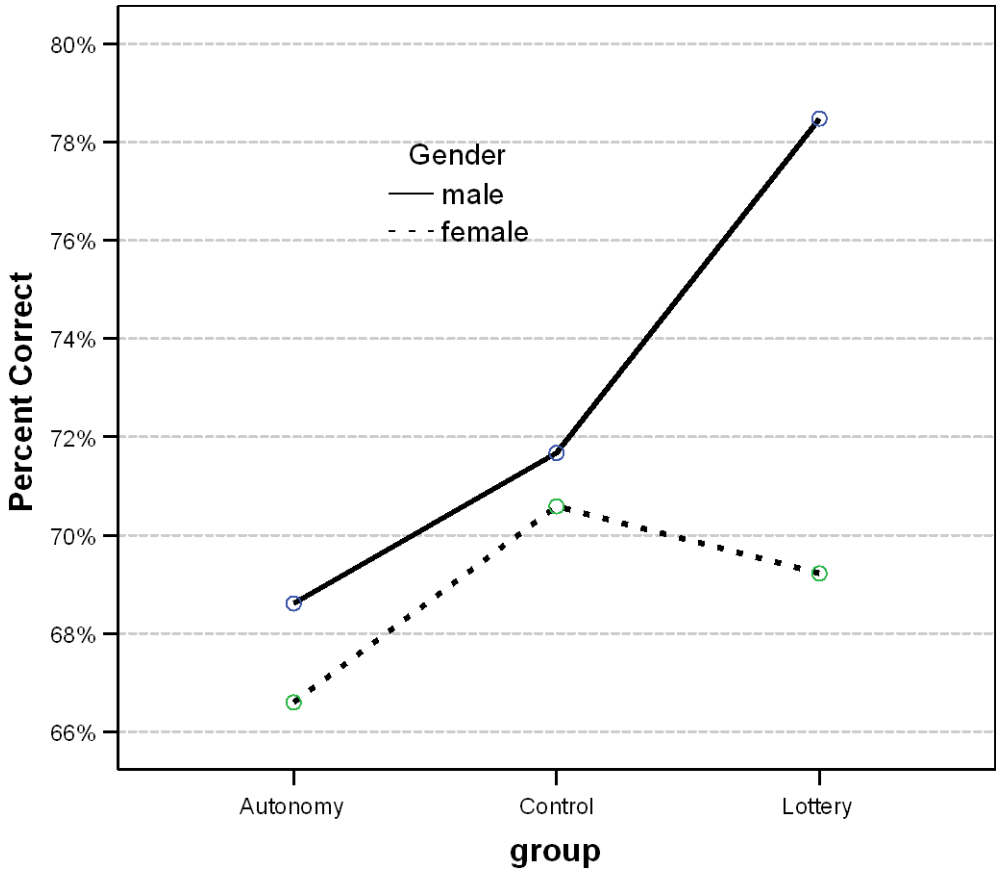
Gender	n	Mean	Std. deviation
Male	87	72.92	1.208
Female	143	68.80	.937

There was a significant interaction between males and females and how they react to the performance contingent extrinsic reward present in the lottery group (Table 5) ($p<.05$). Males in the lottery group had a score 9.26 points higher than females in the lottery group, dramatically higher than the differences in male-female scores for the other two groups (Figure 3). This suggests that males respond to this sort of lottery significantly more positively than females. In other words, the lottery intervention had a significant positive impact on male test performance. The autonomy support intervention had no significant impact on male performance. Neither intervention had a significant impact on female test performance.

Table 5. Adjusted Mean Test Performance Scores by Group and Gender

Group	Gender	n	Mean	Std. deviation
Autonomy	male	25	68.607	2.242
	female	33	66.594	1.530
Control	male	28	71.672	2.106
	female	47	70.584	1.633
Lottery	male	34	78.476	1.915
	female	43	69.220	1.699

Figure 3. Gender interaction with treatment



There will be no difference between the three groups regarding the impact of value/usefulness, choice, interest, and effort on test performance

Hypothesis 3 claimed that there will be no difference between the three groups regarding the impact of value/usefulness, choice, interest, and effort on test performance. More specifically, based on prior research, I expected that effort would likely mediate the impact of gender, prior achievement (ACT scores), interest, value, and choice on test performance. To test the invariance of the model across these three groups and the mediating effects of effort, multi-group path analysis was used using the GROUPS function in EQS (Byrne, 2006; Dunn, Everitt, & Pickles, 1993). All direct path coefficients were constrained to be equal to test for measurement invariance. A single model fit for the three groups was calculated (χ^2 [49] = 20,620.98, $p < .001$; CFI = 0.42; NNFI = 0.188; RMSEA = 1.44).

Given the poor model fit, the Lagrange Multiplier (LM) test was used to determine if any parameters should be added or if any of the constraints should be released. The LM test results indicated that several of the constraints were non-invariant at the p <.01 critical value. As a result, the following constraints were released and allowed to vary between the three groups: gender on percent correct, ACT scores to percent correct, and effort to percent correct. The LM Test also identifies parameters that should be added to improve model fit. The following parameters were added for all three groups: choice on percent correct, interest on percent correct, and value on percent correct. A single model fit for the three groups was calculated (χ^2 [28] = 37.42, $p > .05$; CFI = 0.95; NNFI = 0.897; RMSEA = .07). The model fit for this final model is deemed adequate using fit criteria establish by Hu and Bentler (1999) and Marsh, Hau, and Wen,

(2004). The final path models for each group are included in Appendix F, which includes the respective unstandardized path coefficients.

Table 6. Summary of model fit statistics

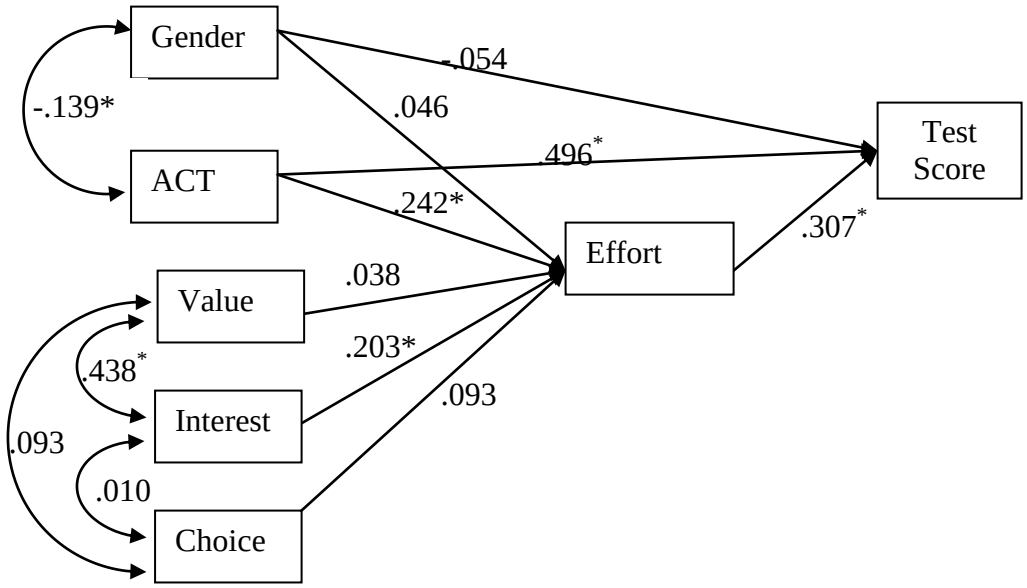
Group	χ^2	df	p	CFI	SRMR	RMSEA	90% C.I. RMSEA
Initial model	20,620.98	49	.000	.424	.158	2.485	2.438; 2.495
Final model	37.419	28	.110	.954	.075	.070	.000; .124

Total variance explained Autonomy = .472
Total variance explained Control = .478
Total variance explained Lottery = .546

Autonomy Group results

The final path model with direct and indirect effects is indicated in Figure 4 and includes standardized path coefficients. ACT had a significant indirect and direct effect on test score. However, gender did not have a significant direct or indirect effect on test score. This is somewhat surprising given past research on gender differences in math performance. Effort had a significant impact on test score ($p<.05$) and mediated the relationship between interest, ACT, and test score. Value and choice had no direct or indirect effect on test score. The results of this final model indicate that effort does partially mediate ACT and interest on test score, and that value and choice have no significant direct or indirect effects on test score.

Figure 4. Final Path Model Results for Autonomy Group (standardized values)

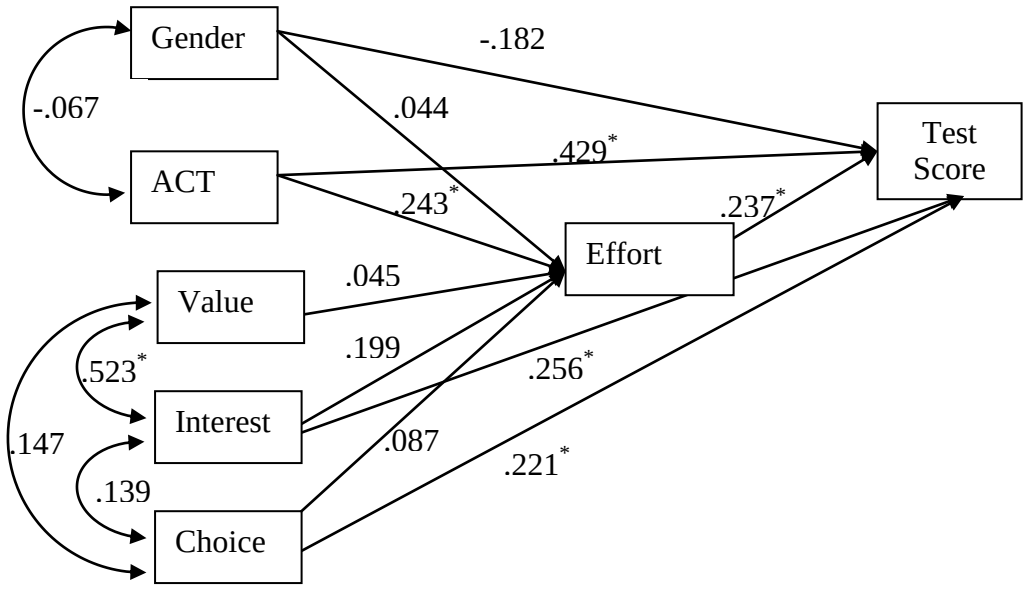


* p<.05

Control Group results

The final path model with direct and indirect effects is indicated in Figure 5 and includes standardized path coefficients. Similar to the autonomy group, ACT had a significant indirect and direct effect on test score (p<.05). Gender did not have a significant direct or indirect effect on math test score. Effort had a significant impact on test score (p<.01), but did not mediate the relationship of any internalization measures (interest, value, choice). Effort was a partial mediator for ACT on test score. Based on the LM test, significant direct effects were found between interest, choice, and test score. The results of this modified model indicate that effort does partially mediate ACT on test score, and that interest and choice have significant direct effect on test score, but no significant indirect effects.

Figure 5. Final Path Model Results for Control Group (standardized values)

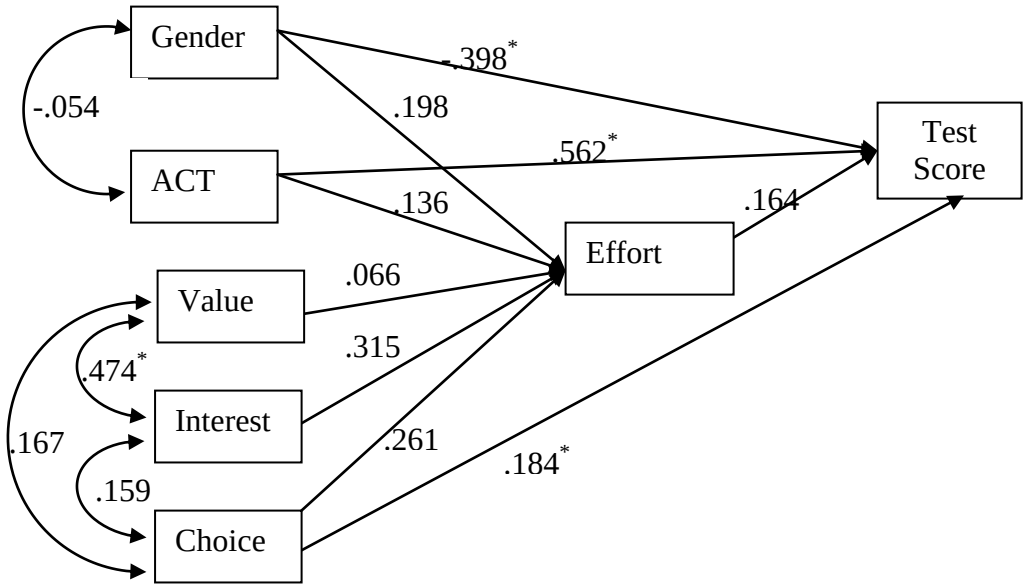


* p<.05

Lottery Group results

The final path model with direct and indirect effects is indicated in Figure 6 and includes standardized path coefficients. ACT and gender had a significant direct effect on test score (p<.05). However, the gender coefficient is negative indicating that males significantly contributed to the variance in test score. Gender did not have a significant indirect effect on math test score. Surprisingly, effort did not significantly mediate any variables on test score. Choice did have a significant direct effect on test score.

Figure 6. Final Path Model Results for Lottery Group (standardized values)



* p<.05

Summary of Path Analysis

The results of the path analysis indicate that effort only partially mediates test performance. The role of effort differed for all three groups. For the autonomy group, effort significantly mediated interest and ACT on test score. However, for the control group, effort only mediated ACT. Effort had no significant mediation role for the lottery group. Also different was the role of the internalization variables: interest, value, and choice. In the autonomy model, none of the internalization measures had a significant direct effect on test score. In the control group however, interest and choice had significant direct effects, while only choice had a direct effect on test score for the lottery group.

Based on these results it appears that the two interventions activate different relationships of key variables during a testing situation. The lottery intervention seemed to significantly favor males during the exam. On the other hand, the autonomy intervention seemed to minimize gender differences, but facilitated the role of interest and effort during the exam.

The amount of variance (R^2) in test scores explained by the data in the three models was .472 for autonomy, .546 for lottery, and .478 for the control group.

CHAPTER V: DISCUSSION

The implications of this research are both practical and theoretical. The practical implications are grounded in the recognition that student motivation to do well on low-stakes tests often is quite low. Low motivation is due to the task being perceived by the student as of low-interest, little value, and no consequence. Given that test performance is a function of both knowledge and motivation, the possibility of low student motivation raises the concern of whether data collected under low stakes conditions are valid measures of student achievement (Eklof, 2006; Wainer, 1993). Institutions want to be confident that their students' test scores accurately represent student knowledge in the academic subject areas tested.

The purpose of this study was to addresses three primary problems: (1) to what extent can student motivation on a relatively low-interest task such as taking a low-stakes test be enhanced by use of a lottery to win money or by use of autonomy support, (2) to what extent can test performance on a relatively low-interest task such as taking a low-stakes test be enhanced on a low-stakes test by use of a lottery to win money or by use of autonomy support, and (3) what is the relationship between test-taking motivation and test performance? To help answer these questions, hypotheses were proposed and tested. Specifically, I tested the efficacy of increasing student motivation to do well on a low-stakes math test in two ways. One was to assign students to a group that was offered a chance to be entered into a lottery for $100 gift card based on their test score. The second group of students received modified test instructions that were intended to facilitate student autonomy and internalization of the low-interest task. The modified test

instructions to the students were intended to create a non-controlling environment, as well as convey a rationale to try your best on the exam. In addition, it was acknowledged that this is probably of low-interest to the students by stating that some students "might not be enthusiastic" about taking the test. It was hypothesized that students in this autonomy-supported group would experience greater levels of internal regulation, thereby enhancing their motivation to do well on the exam. The impact on test performance by offering the chance to win $100 or the use of autonomy support was also tested. Finally, I wanted to find out what would be the impact of the two interventions (financial reward and autonomy support) on a theoretical model of test-taking motivation.

The theoretical purpose of this study was to test the efficacy of enhancing student motivation to do well on a low-stakes test relying solely on the facilitation of autonomy support. Other low-interest studies have successfully used Self-Determination Theory to enhance subject autonomy and internalization of low interest tasks. However, application of this theory has not been applied to a low interest task such as test-taking.

Effectiveness of Autonomy Support and Financial Reward in Enhancing Test-Taking Motivation

The results suggest that providing autonomy support while engaged in a low-interest task did in fact put forth more effort to do well on that task. In this study, students in the autonomy-supported group reported significantly higher levels of test-taking effort compared to those students in the control group. Students who were provided with a rationale reported significantly higher levels of external regulation compared to students who were offered a financial reward and the students offered no

reward or rationale. This result is somewhat perplexing and counter to the expected results. Prior research in low-interest tasks has found that students who were offered a rationale in a non-controlling environment reported higher levels of internal regulation and lower levels of external regulation compared to a control group (Deci et al., 1994, Jang, 2003). The contradictory results from this current study may be due to a moderator effect and is further discussed in the Limitations section.

Students in the lottery group reported significantly higher levels of test-taking effort compared to those students in the control group. This result is consistent with prior studies examining the impact of incentives on test-taking effort (O'Neil, et al., 2005; Wise & DeMars, 2005). However, most previous data do not support the use of providing a rationale or the possibility of financial reward to enhance components of test-taking motivation. For instance, in this study there were no differences between any of the groups with regard to their intrinsic motivation, their interest in doing well on the exam, and their valuing the exam outcomes. This finding of no difference is somewhat surprising given past research suggesting that performance contingent rewards often undermine autonomy (Deci & Ryan, 2000). However, it could be that in this situation students in the lottery group did not perceive the possible financial reward as controlling, and therefore the treatment would not have undermined their interest or value for the activity.

Effectiveness of Autonomy Support and Financial Reward in Improving Test Performance

This study provided clear evidence that students with a chance to be entered into a lottery for $100 based on their test performance outscored students not offered such a

possible reward. On the average, students in this lottery group scored approximately 6 points higher than students who were provided a rationale and approximately 2.5 points higher than students who were offered no rationale or possible financial reward. This finding is consistent with other research investigating the impact of performance-based financial rewards on test performance (O'Neil, et al., 1996). Interestingly, students who were provided autonomy support scored the lowest on the exam. These are the same students who also reported significantly higher levels of effort compared to students in the control group on that same exam. This finding is not consistent with some research which found that students who reported higher levels of effort achieve higher test scores (Cole & Bergin, 2006; Wise & DeMars, 2005). However, this result is consistent with the recent study completed by O'Neil and his colleagues. They report that incentives did in fact increase self-reported effort but that reported test-taking effort was not correlated with test performance (O'Neil et al., 2005).

The main effect of the group comparison should be interpreted cautiously however, because an interaction between groups and gender. This study provided clear evidence that males and females respond quite differently to the possibility of winning a financial reward. The data showed that males responded to the lottery much more positively than did females. Males who were offered a chance at winning the financial reward scored almost 10 points higher on the exam compared to females, even after using ACT scores to control for any differences attributable to prior achievement. In fact, there was no significant difference in test performance between females in any group suggesting that none of the two interventions in this study had a major impact on female test performance in low-stakes situations. These results are similar to other related

studies. A recent study by O'Neil et al. (2005) reported that males scored significantly higher on the TIMMS math exam than females when both were offered financial reward based on performance. It has also been reported that performance contingent rewards have a greater negative impact on girls' reported autonomy (Houlfort, Koestner, Joussemet, Nantel-Vivier, & Lekes, 2002) and that females tend to exhibit lower performance when they find themselves in competitive environments that include men (Gneezy, Niederle, Rustichini, 2003). Though a lottery may improve overall test scores, it may in fact only do so for males.

Impact of Autonomy Support and Lottery on Theoretical Understanding of Test-Taking Motivation

Based on Self-Determination Theory it was proposed that indicators of internalization (value, interest, and choice) would influence test-taking effort that would then influence test performance. Results from this analysis are mixed. The results suggest that important variables in predicting test scores vary somewhat depending on the situation. For example, valuing the activity and perceiving some choice or control over participation are important predictors of test-taking effort and performance for students in the autonomy group. This result is not surprising and is consistent with other recent research investigating influence of internalization on task-specific effort (Jang, 2003). The three most important predictors of test performance for these students were ACT scores, choice, and value mediated by effort (in that order). Overall, almost 50% of the variance in test performance was explained by a path model of test-taking performance.

In contrast, valuing the activity was a less important predictor of test scores for those students offered the chance to win a financial reward based on their performance. For these students, experiencing control or choice in their participation only minimally influenced their test scores. This result is also consistent with Self Determination Theory. As the theory predicted, students in this group did not internalize the low-interest activity and thus did not relate their level of internalization with their test-taking effort. Instead, the two most important predictors of test performance for these students were gender and ACT scores. Overall, approximately 44% of the variance in test performance was explained by the path model.

Factors influencing test performance were quite different for those students in the control group who received neither autonomy support nor possibility of financial reward. Unlike students who received autonomy support or possibility of financial reward, for control students, the role of value and its influence on test-taking effort was quite minimal. In fact, the direct effect of value on test performance was moderately negative for this group in stark contrast to the moderately positive influence that value had on test performance for the other two groups. Another difference is the role that interest played in predicting test-taking effort and test performance. For the students in the control group, interest was a relatively important predictor especially when compared to the minimal role that interest had on effort and test performance for students in the other two groups. These results suggest that in the absence of autonomy support or possible financial reward, task-specific interest plays a more prominent role in predicting effort and performance and that valuing the activity has a minimal impact on effort but a significant negative impact on test scores. Conversely, in the presence of autonomy

support or possible financial reward, the role of interest is clearly undermined while the role of value is greatly enhanced. These results are not fully understood in the context of Self-Determination Theory, which would predict that interest in the activity and valuing the activity would be enhanced by autonomy support (Deci & Ryan, 2002). Clearly, more research needs to be done to further investigate this phenomenon.

Implications

The present study adds to the current and growing body of literature on test-taking motivation in a number of ways. First this study reaffirmed the important role that effort plays in test performance. For the autonomy and control groups, test taking effort had a significant impact on test performance. The standardized path coefficients for the autonomy and control groups were .307 and .234, respectively. Many other studies have also reported the important role of effort in test performance (Wise & DeMars, 2005). However, this study also found evidence to support the conclusion by O'Neil et al. (2005) that increased effort does not necessarily correlate with increase test performance. Similar to that study, this study found that students who received autonomy support did in fact report higher levels of test effort, but scored significantly lower than students in the lottery group. Conversely, effort played a non-significant role in predicting test performance for the lottery group, even though they reported the highest levels of effort. This study was also important because it re-examined the use of a lottery to increase motivation and test-scores. The results demonstrate that scores can be improved with the use of a performance-based lottery system, but only for male test-takers. Though this study did not collect data on perceptions of competitiveness, other research has found that

males tend perform better under competitive situations compared to females (Gallagher, 1998). It could be that a lottery system that pits students against each other for the prizes is perceived as competitive by males, which may facilitate their motivation. It is important to keep in mind though that data from this study found no significant difference in test scores between female students in any group, suggesting that the autonomy support and the lottery system had no impact on female test performance. There is some evidence from this study though that the autonomy intervention may have helped to minimize gender differences in test performance. The impact of gender on test scores for that group was quite small, compared to the control and lottery groups. However, given these results and until such time as other research can investigate this issue, it is not recommended that schools try to increase test scores by establishing a lottery system.

The implication of these results is that educators, psychometricians, administrators, and others interested in assessing student outcomes need to recognize that test-taking effort is an important predictor of test performance and that we can not assume that all students give maximum effort when taking a low-stakes test. However, this study also indicates that the relationship of effort to performance is not fully understood. More research needs to be conducted to better understand how to encourage maximal effort and performance from students during exam situations.

Another implication of this study is that use of a lottery system will likely increase test performance, but only for males. Schools that offer such rewards are doing so with the best of intentions, but they may actually be undermining female performance on these exams. This study and one conducted by O'Neil et al, (2005) both found that female test performance was undermined by the use of financial rewards. Until more

study can be conducted investigating this issue, schools may want to reconsider offering lotteries and other financial rewards to their students.

The study was also important in that it was the first to tests the efficacy of autonomy support in raising test-taking motivation and performance. The results however, were disappointing. Though self-reported effort was significantly enhanced for those students who received the autonomy support instruction, their test scores were not significantly different than of those students in the control group. This study produced no evidence that providing autonomy support through a script like the one used will positively affect test performance. This may imply that an autonomy supportive intervention may not be generally efficacious at enhancing performance on low-interest tasks. Certainly the results of this study do not support the efficacy of this type of intervention during the administration of a low-stakes test. However, this is only one study. These results should be seen as an impetus to conduct more research on autonomy support, and not the abandonment of this type of intervention and use in low-stakes testing situations. It could be that the specific wording in the instructions used in the autonomy intervention was not effective in enhancing the student's autonomy. Modifying the instructions to better enhance autonomy may produce different results.

This study also highlights how the test-taking situation can influence the processes that contribute to test performance. This study found that variables that are important in predicting test-taking effort and test performance were easily manipulated during test administration. Important predictors of effort and test performance varied between all three groups. The implications are that we need to recognize the fragile nature of this situation. Ideally, standardized tests reduce this type of measurement error

by "standardizing" administration procedures and protocols. However, many aspects of the testing situation may not be consistent across all test administrations. Care needs to be taken to assure that test proctors, educators, and school administrators are not inappropriately or accidentally influencing the testing environment, especially given that the results of some interventions may well undermine certain student scores rather then enhance them.

Limitations

The primary limitation of this study is the concern regarding ecological validity (Brewer, 2000). Did the artificiality of this experiment truly replicate the real-world environment students find themselves in when asked to complete a low-stakes exam? For instance, did paying all the students $15 to participate in the study affect the students' effort or test performance? Would these results replicate if conducted in a school setting in more authentic conditions? These questions point to the need for a multi-faceted approach to understanding test-taking motivation. No one study provides all the answers; however, this study provides one more piece of the puzzle that constitutes test-taking motivation and performance.

Readers may also be concerned because only college age students from one university were included in the study. Therefore, these results may not be representative of other college students or students in other grade levels. The generalizability of these results to younger students is not fully known. For instance, the use of a lottery system may not be effective with young school age children who may not fully understand how their odds or probability of winning money are affected by their test performance. The

use of autonomy support may be even more difficult to convey to high school students who constitute a broader spectrum of academic abilities, some of whom have no desire for higher education.

Some may also be concerned with construct validity of measurement, especially with the modified internalization and regulation scales. Though the scales were adapted from known psychological instruments with established psychometric properties, the scales used in this study were modified. With no criterion variables to relate the new modified scales to, it is not fully known if the underlying latent constructs were changed (Anastasi & Urbina, 1997). What is known is that the data collected in this study did adequately factor analyze into the intended factors. Therefore, the data collected from these scales were deemed adequate for further analysis in the study.

There is also the possibility that the autonomy manipulation may not have been effective. The results of this study were certainly disappointing in this regard. Though the autonomy intervention was successful at enhancing test-taking effort, it was not successful in enhancing test performance as expected. Would the results have been different if a different proctor administered the autonomy supportive instructions? Could the script have been modified in such a way to have a stronger effect on test performance? If the intervention was the script, should the questionnaire containing the measures of internalization and autonomy support been completed immediately after the script and prior to taking the exam? Other similar SDT studies have done so. However, it was felt that to get accurate measures of test-taking effort and motivation, the questionnaire should be given to the test-takers after completing the exam, not prior to

their participation in the exam. This may have caused the measures of internalization to be confounded by the experience of taking a math test.

It is possible that the lack of performance feedback may have undermined the reported effort and motivation of this group. The script for the autonomy group implies that each student will receive feedback regarding his/her performance on the exam. However, students completed the questionnaire immediately after the exam and before they had time to realize that they were not going to find out their test score. Therefore, the effect of lack of feedback was likely negligible. However, the use of performance feedback is an important motivator in educational settings. The use of this type of treatment to improve test performance should certainly be explored. Future research should try to replicate and improve on the autonomy supportive procedures used in this study.

Future Research

With the limitations acknowledged above, it is also important to recognize the contribution of this research to future research. Certainly one recommendation for future research is to further investigate the effect of possible financial reward on the test performance for groups of students. This study found that the possibility of financial reward for performance on a low-stakes math test significantly improved male test performance, while having no affect on female performance. Future research may want to try to replicate testing this important finding, as well as investigate the generalizability of this result to other groups and/or other subject areas. For instance, would this study have had the same result if a English composition and literature test were used rather than

a math test? It is well established that females often out perform males on such tests (Halpern, 2000) and therefore interventions to enhance motivation may operate differently depending on gender as was the case with the use of a lottery on a math test in this study.

This study investigated only two possible interventions to improve test-taking effort and performance. Other interventions may well prove effective. One possible intervention that should be tested is to provide immediate and/or delayed performance feedback (Wise, 2004). It is well-established that good classroom practice for teachers includes providing timely and accurate feedback to students regarding their performance (Ormrod, 2005). Future research may want to investigate the impact of providing immediate feedback to students regarding their performance on a low-stakes exam.

Future research should continue to focus on the use of autonomy support to enhance test-taking motivation for low stakes tests. Though this study found that autonomy support improved self-reported test-taking effort, but not test scores, many low-interest studies grounded in Self-Determination Theory have been successfully conducted over the years with promising results (Deci et al., 1994; Deci & Ryan, 2002; Jang, 2003; Joussemet et al., 2004). The efficacy of providing a meaningful rationale to individuals engaged in low-interest activities is well-established. Future research should continue investigate the effectiveness of autonomy support in low-stakes testing situations.

This study along with others confirmed the important role that effort plays a predicting test scores. What is still not clear, however, is what exactly influences test-taking effort and how those influences may vary by group or testing situation.

REFERENCES

Anastasi, A. & Urbina, S. (1997). *Psychological Testing*. Upper Saddle River, NJ: Prentice Hall.

Banta, T. W. (2002). *Building a scholarship of assessment*. San Francisco: Jossey-Bass.

Baines, L. A. & Stanley, G. K. (2003). Disengagement and loathing in high school. *Educational Horizons, 81*, 165-168

Baumert, J. & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education, 16*, 441-462.

Bergin, D. A. (1999). Influences on classroom interest. *Educational Psychologist, 34*, 87-98.

Brewer, M. (2000). Research Design and Issues of Validity. In Reis, H. and Judd, C. (eds) *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press.

Cole, J. S. & Bergin, D. A. (2006, April). *Predicting student achievement for low stakes tests with effort and task value*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Cole, J. S. & Osterlind, S. J. (2005, April). *Effects of test consequence with college students*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada

DeCesare, D. (2002). How high are the stakes in high-stakes testing? *Principal, 81*, 3, 10-12.

Deci (1975). *Intrinsic motivation*. New York, NY: Plenum.

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality, 62,* 119-142.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research, 7,* 1-27.

Deci, E. L., & Ryan, R. M. (1985*). Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Deci, E. L. & Ryan, R. M. (2002). *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.

DeMars, C. E. (2000). Test stakes and item format interactions. *Applied Measurement in Education, 13,* 55-77.

Eklof, H. (2006). Development and validation of scores from an instrument measuring test-taking motivation. *Educational and Psychological Measurement, 66,* 643-656.

Erwin, T. D. & Wise, S. L. (2002). A scholar-practitioner model for assessment. In T. W. Banta, T. W. (Ed), *Building a scholarship of assessment (pp. 67-81).* San Francisco: Jossey-Bass.

Gallagher, Ann (1998). Gender and antecedents of performance in math testing. *Teachers College Record, 2,* 297-315.

Gneezy, U., Niederle, M., & Rustichine, A. (2003). Performance in competitive Environments: Gender differences. *Quarterly Journal of Economics, 118,* 1049-1074

Goertz, M & Duffy, M. (2003). Mapping the landscape of high-stakes testing and accountability programs. *Theory into Practice, 42*, 4-11.

Good & Brophy (2003). *Looking into classrooms*. Boston, MA: Allyn and Bacon.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Hancock, G. R. & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural Equation Modeling: Present and Future- A Festschrift in honor of Karl Jöreskog* (pp. 195-216). Lincolnwood, IL: Scientific Software International, Inc.

The Higher Learning Commission. (2002). *Addendum to the handbook of accreditation* (2nd ed.). Chicago, IL: North Central Association of Colleges and Schools.

Houlfort, N., Koestner, R., Joussemet, A. N., & Lekes, N. (2002). The impact of performance-contingent rewards on perceived autonomy and competence. *Motivation and Emotion, 26,* 279-295

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Jang, H. (2003). *Providing a rationale to engage students in an uninteresting activity: A test of multiple models.* Unpublished doctoral dissertation, University of Iowa.

Joussemet, M., Koestner, R., Lekes, N., & Houlfort, N. (2004). Introducing uninteresting tasks to children: A comparison of the effects of rewards and autonomy support. *Journal of Personality, 72,* 139-166

Karmos, A. H. & Karmos, J. S. (1984). Attitudes toward standardized achievement tests and their relation to achievement test performance. *Measurement and Evaluation in Counseling and Development, 17*, 55-66.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.

Kober, N., Zabala, D., Chudowsky, N., Chudowsky, V., Gayler, K. & McMurrer, J. (2006). *State high school exit exams: A challenging year*. Washington DC: Center on Education Policy.

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning, and development. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp 3-25). Hillsdale, NJ: Erlbaum.

Linn, R. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*(7), 3-13.

Linn, R. (1990). Admissions testing: Recommended uses, validity, differential prediction, and coaching. *Applied Measurement in Education, 3*, 297-318.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*, 320-341.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport, 60*, 48-58.

Missouri Department of Elementary and Secondary Education (2006). *Missouri Assessment Program.* Jefferson City, MO.

Missouri Department of Elementary and Secondary Education (1997). *Missouri Mastery and Achievement tests: Guide to score interpretation and use*. Jefferson City, MO.

Napoli, A. R. & Raymond, L. A. (2004). How reliable are our assessment data?: A comparison of the reliability of data produced in graded and ungraded conditions. *Research in Higher Education, 45*, 921-929.

No Child Left Behind Act of 2001, Pub. L. No. 107-110.

Olsen, S. A. & Wilson, K. (1991). *A follow-up of suspect sophomore scores on the COMP test.* Kirksville, MO: Northeast Missouri State University. (ERIC Document Reproduction Service NO. ED 339735).

O'Neil, H. F., Sugrue, B., & Baker, E. L. (1996). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment, 3*, 135-157.

O'Neil, H.F., Jr., Abedi, J., Miyoshi, J., and Mastergeorge, A. (2005). Monetary Incentives for Low-Stakes Tests. *Educational Assessment, 3*, 185-208.

Ormrod, J. E. (2005). *Educational psychology: Developing learners*, 5th Edition. Englewood Cliffs, NJ: Prentice-Hall.

Palomba, C. A. & Banta, T. W. (1999). *Assessment essentials: Planning implementing, and improving assessment in higher education*. San Francisco: Jossey-Bass.

Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher, 20*(5), 12-20.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications.* Englewood Cliffs, NJ: Prentice-Hall.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* National Center for Research to Improve Postsecondary Teaching and Learning. University of Michigan, School of Education, Technical Report No. 91-B-004.

Reeve, J. (2001). *Understanding motivation and emotion*. New York, NY: Wiley.

Reeve, J., Jang, H., Hardre, P., & Omura, M. (2002). Providing a rationale in an autonomy-supportive way as a strategy to motivate others during an uninteresting activity. *Motivation and Emotion, 26*, 183-207.

Reeve, J., Nix., G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology, 2*, 375-392.

Renninger, K. A. & Hidi, S. (2002). Student interest and achievement: Developmental issues raised by a case study. In A. Wigfield, & J. S. Eccles (Eds.), *Development of achievement motivation* (pp 175-191). San Diego, CA: Academic Press.

Ryan, R. N. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*, 450-461.

Ryan, R. M. & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54-67.

Ryan, R. M. & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68-78.

Sanders, T. (2003). Clearing the NCLB hurdle. *American School Board Journal, 190(*9), 26-28.

Smith, L. F. & Smith, J. K. (2004). The influence of test consequences on national examinations. *North American Journal of Psychology, 5*, 13-26.

Steeves, K. A., Hodgson, J., & Peterson, P. (2002). Are we measuring student success with high-stakes testing? *The Educational Forum, 66*, 228-235.

Sundre, D. L. & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29*, 6-26.

Tabachnick, B. G. & Fidell, L. S. (2000). Using multivariate statistics. Boston, MA: Allyn & Bacon.

Thomas, R. M. (2005). *High stakes testing: Coping with collateral damage*. Mahwah, NJ: Lawrence Erlbaum.

Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement, 30*, 1-21.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25,* 68-81.

Wigfield, A., & Eccles, J. S. (2002). *Development of achievement motivation*. San Diego, CA: Academic Press.

Wise, S. L. & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10,* 1-17.

Wise, S. L. (2006). An investigation of the differential effort received by items on a low-stakes computer-based test. *Applied Measurement in Education, 19,* 95-114.

Wise, V. L. (2004**).** *The effects of the promise of test feedback on examinee performance and motivation under low-stakes testing conditions*. Unpublished doctoral dissertation, University of Nebraska, Lincoln.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8,* 227-242.

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test, motivation, and mentally taxing items. *Applied Measurement in Education, 8,* 341-351

Appendix A: INFORMED CONSENT LETTER

April, 2006

Dear MU student,

Thank you for your interest in participating in this study. This study is being conducted by James Cole from the Assessment Resource Center and Dr. David Bergin from the Department of Educational, School, and Counseling Psychology at the University of Missouri-Columbia. In this study we will ask you to complete a short standardized exam and a questionnaire concerning your attitudes and experiences. Your participation in this project is voluntary, and you do not have to answer any questions you do not want to answer.

The results of this project may be published, but only group data will be reported. Your confidentiality will be safeguarded. This project is not expected to involve risks greater than those ordinarily encountered in daily life. Although it is not possible to identify all potential risks in the study, all reasonable safeguards have been taken to minimize any potential risks. If you decide to withdraw during the study, you will not be penalized in any way. You must be at least 18 to take part in the study.

If you have any questions or comments regarding the administration of this survey, please raise your hand and the test proctor will assist you. If you have further questions regarding this research, please contact James Cole at the Assessment Resource Center at 573-882-2963 or by email at: colejs@missouri.edu. If you have questions regarding your rights as a research participant, you may contact the UMC Institutional Review Board at 573-882-9585 or by email: umcresearchcirb@missouri.edu. An additional copy of this informed consent for you to keep is provided for your records.

Thank you very much for helping with this important study.

Sincerely,

James Cole
Sr. Coordinator

David Bergin, Ph.D.
Associate Professor

- -

Please sign below to indicate that you understand and give your consent to participate in the study. Your signature does not constitute a waiver of any legal rights.

____

Signed name Date

Appendix B: SURVEY

Please provide your student ID number: _____

The following items concern your experience with the task. Please answer all items. For each item, please indicate how true the statement is for you, using the following scale as a guide:

not at all true somewhat true very true

Why would I try to do well on this test?

1. because I want the experimenter to think I am a good student.	1	2	3	4	5	6	7
2. because I'm supposed to	1	2	3	4	5	6	7
3. because that's what the experimenter says I should do	1	2	3	4	5	6	7
4. because I will feel bad about myself if I don't try to do well	1	2	3	4	5	6	7
5. because I pressure myself to do well on tests	1	2	3	4	5	6	7
6. because I know that I should	1	2	3	4	5	6	7
7. because I want to check my understanding of math	1	2	3	4	5	6	7
8. because its important to me to do well at math	1	2	3	4	5	6	7
9. because I value contributing to research studies	1	2	3	4	5	6	7
10. because its challenging to take tests	1	2	3	4	5	6	7
11. because I enjoy taking tests	1	2	3	4	5	6	7
12. because its fun	1	2	3	4	5	6	7

Please also answer these questions.

not at all true somewhat true very true

13. I believe that doing this activity could be of some value for me.	1	2	3	4	5	6	7
14. I believe I had some choice about doing this activity.	1	2	3	4	5	6	7
15. While I was doing this activity, I was thinking about how much I enjoyed it.	1	2	3	4	5	6	7
16. I believe that doing this activity is useful for improved concentration.	1	2	3	4	5	6	7
17. This activity was fun to do.	1	2	3	4	5	6	7
18. I think this activity is important for my improvement.	1	2	3	4	5	6	7
19. I enjoyed doing this activity very much.	1	2	3	4	5	6	7
20. I really did not have a choice about doing this activity.	1	2	3	4	5	6	7
21. I did this activity because I wanted to.	1	2	3	4	5	6	7
22. I think this is an important activity.	1	2	3	4	5	6	7
23. I felt like I was enjoying the activity while I was doing it.	1	2	3	4	5	6	7

Continued on next page

	not at all true		somewhat		true		very true
24. I thought this was a very boring activity.	1	2	3	4	5	6	7
25. It is possible that this activity could improve my studying habits.	1	2	3	4	5	6	7
26. I felt like I had no choice but to do this activity.	1	2	3	4	5	6	7
27. I thought this was a very interesting activity.	1	2	3	4	5	6	7
28. I am willing to do this activity again because I think it is somewhat useful.	1	2	3	4	5	6	7
29. I would describe this activity as very enjoyable.	1	2	3	4	5	6	7
30. I felt like I had to do this activity.	1	2	3	4	5	6	7
31. I believe doing this activity could be somewhat beneficial for me.	1	2	3	4	5	6	7
32. I did this activity because I had to.	1	2	3	4	5	6	7
33. I believe doing this activity could help me do better in school.	1	2	3	4	5	6	7
34. While doing this activity I felt like I had a choice.	1	2	3	4	5	6	7
35. I would describe this activity as very fun.	1	2	3	4	5	6	7
36. I felt like it was not my own choice to do this activity.	1	2	3	4	5	6	7
37. I would be willing to do this activity again because it has some value for me.	1	2	3	4	5	6	7

38. On a scale of 0 to 100% where 100% means you put forth maximum effort, approximately how hard did you try on this test?

No effort → Maximum effort

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
O	O	O	O	O	O	O	O	O	O	O

39. Please indicate your gender
__ male
__ female

40. What was your overall ACT score? _____

40a. What was your ACT math score? _____

Continued on next page

41. What is your current GPA at MU? _____

42. How would you best describe yourself?

__ Caucasian
__ Black/African American
__ Hispanic
__ Asian
__ American Indian
__ Mixed race/ethnicity
__ Other

43. What is your major? _____

44. What year were you born? _____

45. Have you complete a *college-level* mathematics class at MU or any other institution (this includes math AP classes)?

___ Yes ___ No

45a. ***If yes***, please check the highest college-level math class you have completed (check only one)

___ Intermediate algebra (MAT 110)
___ College algebra (MAT 1100 or 1120)
___ College trigonometry (MAT 1140)
___ Pre-calculus (MAT 1160)
___ Math course beyond pre-calculus
___ Other. Please describe: _____

45b. What was your grade for this highest level class completed (e.g., A-, B, etc): ___

==========

Scales with corresponding item numbers

External Regulation: q1, q2, q3

Identified Regulation: q4, q5, q6

Intrinsic Regulation: q10, q11, q12

Interest: q15, q17, q19, q23, q24, q27, q29, q35

Value: q13, q16, q18, q22, q25, q28, q31, q33, q37

Choice: q14, q20, q21, q26, q30, q32, q34, q36

Effort: q38

Appendix C: DEMOGRAPHIC CHARACTERISTICS

Demographic Characteristics of Participants

	Autonomy	Control	Lottery
Age			
M	19.8	19.6	20.2
SD	0.701	0.747	4.083
N	78	72	79
Gender			
Male	34.6%	38.0%	42.2%
Female	65.4%	62.0%	57.8%
College class			
Freshman	70.7%	78.8%	78.6%
Sophomore	29.3%	21.3%	21.4%
Ethnicity			
Caucasian	77.8%	77.2%	77.5%
Black	9.9%	11.4%	12.5%
Hispanic	1.2%	1.3%	0.0%
Asian	2.5%	2.5%	5.0%
American Indian	1.2%	2.5%	0.0%
Mixed	2.5%	5.1%	2.5%
Other	4.9%	0.0%	2.5%
Total percent	100%	100%	100%
Total N	81	79	80

Appendix D: FACTOR LOADINGS

Pattern Matrix: Regulation scales

	Factor Loadings			
	1	2	3	4
Q1 external	.311		.374	
Q2 external			.769	
Q3external			.760	
Q4 introjected	.858			
Q5 introjected	.739			
Q6 introjected	.478		.346	
Q7 identified				.816
Q8 identified				.682
Q9 identified				
Q10 intrinsic		.327		.395
Q11 intrinsic		.918		
Q12 intrinsic		.790		

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.
Note: only coefficients over .35 are displayed

Pattern Matrix: Activity Perception Questionnaire Scales

Factor Loadings

	1	2	3
Q13 value			.422
Q26 value			.603
Q18 value			.919
Q22 value			.311
Q25 value			.771
Q28 value			
Q31 value			.772
Q33 value			.547
Q37 value			
Q14 choice		.405	
Q20 choice		.693	
Q21 choice		.462	
Q26 choice		.795	
Q30 choice		.783	
Q32 choice		.825	
Q34 choice		.325	
Q36 choice		.529	
Q15 interest	.820		
Q17 interest	.958		
Q19 interest	.900		
Q23 interest	.881		
Q24 interest	.576		
Q27 interest	.645		
Q29 interest	.902		
Q35 interest	.883		

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.
Note: only coefficients over .35 are displayed

Appendix E: INTERCORRELATIONS

Intercorrelations

All Groups (n= 246)

	1	2	3	4	5	6	7	8	9	10
1. Percent correct	1	.541*	.343*	.251*	.086	.257*	.097	.179	-.014	.152
2. ACT		1	.234*	.322*	.021	.169*	.114	.141*	-.051	-.052
3. Test-taking Effort			1	.097	.260*	.270*	.236*	.196*	.211*	.140*
4. Intrinsic motivation				1	.294*	.113	.060	.732*	.353*	-.022
5. Identified motivation					1	.476*	.228*	.378*	.476*	.101
6. Introjected motivation						1	.448*	.179*	.306*	.114
7. External motivation							1	.112	.165*	.033
8. Interest								1	.490*	.082
9. Value									1	.134*
10. Choice										1

* $p < .05$

Autonomy Group (n= 82)

	1	2	3	4	5	6	7	8	9	10
1. Percent correct	1	.620*	.311*	.356*	.208	.427*	.256*	.218	.066	.123
2. ACT		1	.152	.476*	.198	.327*	.034	.212	-.005	-.165
3. Test-taking Effort			1	.066	.507*	.290*	.243*	.099	.318*	.144
4. Intrinsic motivation				1	.350*	.279*	-.034	.727*	.295*	-.115
5. Identified motivation					1	.517*	.205*	.474*	.421*	.134
6. Introjected motivation						1	.397*	.203	.309*	.110
7. External motivation							1	-.024	.087	.019
8. Interest								1	.444*	.008
9. Value									1	.109
10. Choice										1

* $p < .05$

Control Group (n= 80)

	1	2	3	4	5	6	7	8	9	10
1. Percent correct	1	.533*	.360*	.270*	.116	.249*	.003	.249*	-.144	.218
2. ACT		1	.252*	.314*	.134	.129	.089	.199	-.040	-.052
3. Test-taking Effort			1	.114	.166	.239*	.298*	.267*	.077*	.115
4. Intrinsic motivation				1	.348*	.005	.048	.719*	.406*	.082
5. Identified motivation					1	.348*	.150	.373*	.484*	-.091
6. Introjected motivation						1	.485*	.061	.186	-.009
7. External motivation							1	.071	.190	-.138
8. Interest								1	.544*	.152
9. Value									1	.136*
10. Choice										1

* $p < .05$

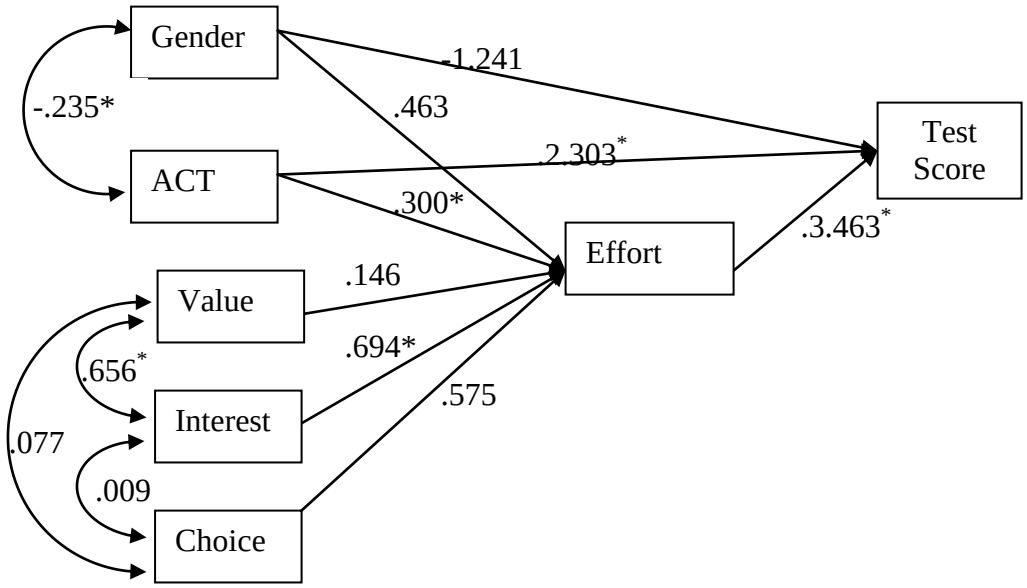
Lottery Group (n= 84)

	1	2	3	4	5	6	7	8	9	10
1. Percent correct	1	.500*	.371*	.130	-.073	-.043	.064	.084	.024	.134
2. ACT		1	.262*	.179	-.290*	.002	.133	.012	-.113	.022
3. Test-taking Effort			1	.134	.069	.246*	.079	.244*	.265*	.148
4. Intrinsic motivation				1	.186	.067	.171	.758*	.344*	-.050
5. Identified motivation					1	.579*	.333*	.301*	.533*	.244*
6. Introjected motivation						1	.498*	.267*	.464*	.254*
7. External motivation							1	.251*	.229*	.160
8. Interest								1	.479*	.076
9. Value									1	.151*
10. Choice										1

* $p < .05$

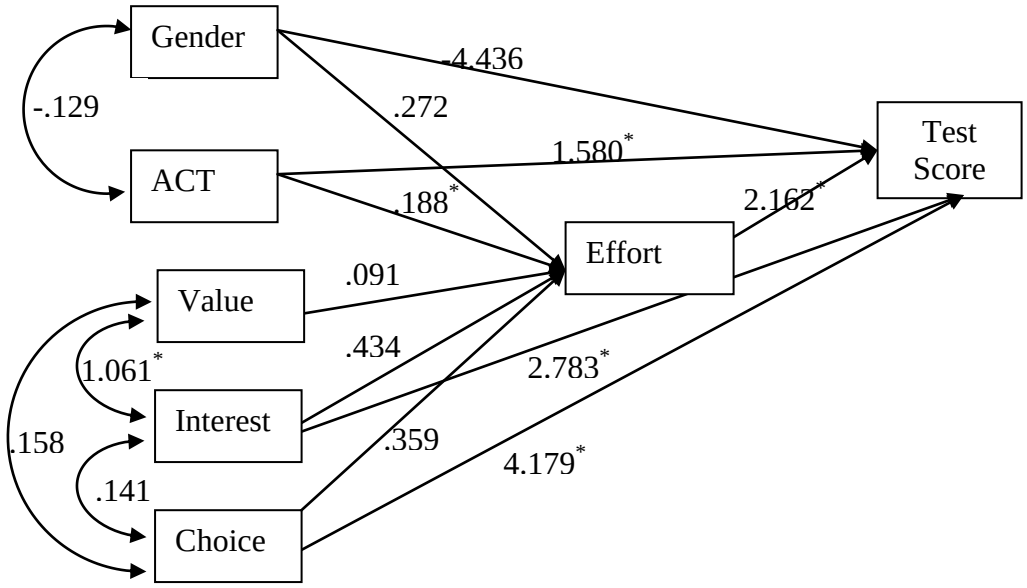
Appendix F: PATH MODEL RESULTS (UNSTANDARDIZED)

Final Path Model Results for Autonomy Group (unstandardized values)



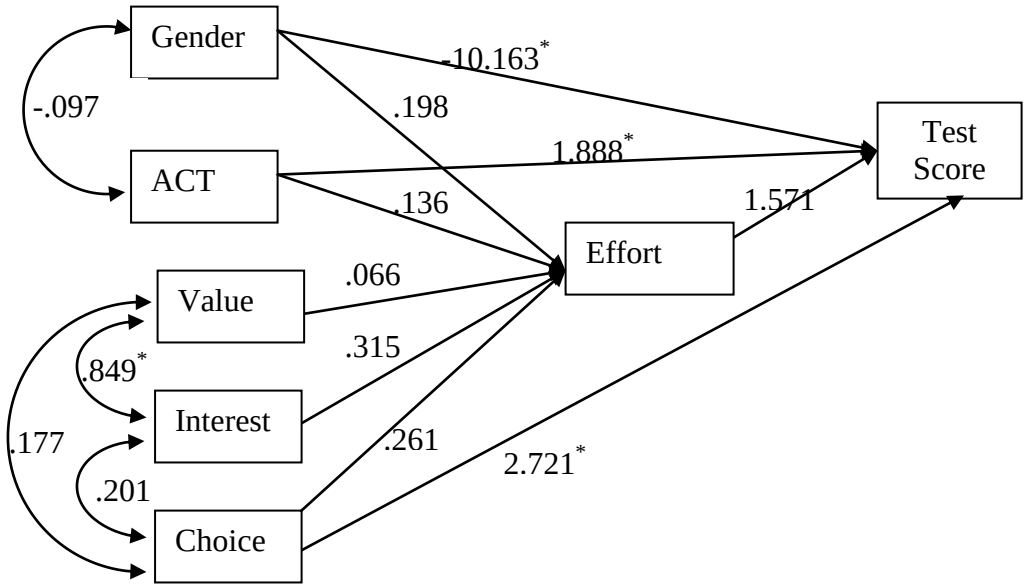
* p<.05

Final Path Model Results for Control Group (unstandardized values)



* p<.05

Final Path Model Results for Lottery Group (unstandardized values)



* p<.05

VITA

James Cole was born in Wrentham, Massachusetts. He earned is A.A. from Sterling College, Vermont and his B.A from Burlington College, Vermont. He continued his education at Texas A&M University where he earned his M.S. in recreation, parks, and tourism sciences.

While at Texas A&M, he met his future wife, Shu. They were married in 1998 and currently have two children. Jim now is employed at Indiana University, Bloomington at the Center for Postsecondary Research as a research analyst.