An Attributional Analysis of Standardized Testing and Outcome Expectancy:
The Results Are In

A thesis submitted to the Miami University
Honors Program in partial fulfillment of the
requirements for University Honors with Distinction

by

James J. Doolittle

May, 2007
Oxford, Ohio
ABSTRACT

An Attributional Analysis of Standardized Testing and Outcome Expectancy: The Results Are In

By
James J. Doolittle

The current study examines a crucial aspect of standardized testing—the presentation of results to students. Participants of this study were college-aged and recruited from a Midwestern mid-sized university. They were asked to complete 41 math and verbal SAT questions, which were later scored and presented to the students in a criterion- or norm-referenced manner. The effects these two types of feedback have on the participants’ attributions (to effort, ability, luck, and task difficulty) and their future expectancy were examined. The study found that the method of reporting does not affect what the participants attributed their success or failure outcome to, or their future expectancy. The limitations of this study and suggestions for future directions are discussed.
Acknowledgements

There are several people and entities that supported the completion of this thesis and therefore deserve acknowledgement and my appreciation. My mentor, Dr. Rose Marie Ward deserves the greatest recognition for her continued dedication not only to the completion of this research paper, but also to my personal, academic, and overall undergraduate growth. I am also greatly appreciative of the Miami University Honors & Scholars Program for their extremely generous financial support that aided in the funding and dissemination of my research, especially in funding a trip to the National Conference on Undergraduate Research at Dominican University. I would also like to thank the Miami University College of Arts and Sciences and the Miami University Office for the Advancement of Research and Scholarship (OARS) for their valuable funding. OARS also deserves recognition for processing in a very timely and efficient manner the many Institutional Review Board protocol revisions I submitted throughout the research process. I also owe a debt of gratitude to my thesis readers, Dr. John Ward and Dr. Kurt Hugenberg, for taking their time to read over my thesis, offering helpful suggestions, and aiding in the completion of the project.

I dedicate this paper to my parents who have afforded me exceptional educational opportunities and have provided unconditional support in all of my endeavors.
## Table of Contents

- Introduction 1
- Health and Academic Performance 3
- The Many Dimensions of Attribution Theory 5
- Self-Serving Bias and Other Potential Variables of Interest
  - Self-Serving Bias 12
  - Gender 14
  - Race/Ethnicity 17
  - Age 19
- Attribution-Expectancy Link 22
- Task- and Ego-Involving Situations 24
- Criterion- and Norm-Referenced Feedback 27
- Overview of the Literature and Hypotheses 31
- Method
  - Participants 37
  - Design and Materials 37
  - Procedure 39
- Results
  - BMI and Test Performance 41
  - Attributions and Future Expectancy Following Success 42
  - Attributions and Future Expectancy Following Failure 43
  - Investigational Analyses 43
- Discussion
  - Expectancy-Valence Theory 46
  - Causal Attributions 48
  - Sample Size 50
  - Learned Helpless versus Mastery-Oriented Dispositions 51
- Conclusions 52
- References 54
- Appendices 64
Tables and Figures

Tables

Table 1. Causes of Success and Failure and their Corresponding Dimensions

Table 2. Criterion- and Norm-Referenced Test Differences

Table 3. Bivariate Correlations among SAT Scores

Table 4. Success and Failure Groups under Criterion- and Norm-Referenced Conditions

Figures

Figure 1. Breakdown by Weight Category (BMI)
An Attributional Analysis of Standardized Testing and Outcome Expectancy:

The Results Are In

One of the most amazing and fundamental features of human beings is the strong desire to understand and explain what is going on in the world. Whether figuring out why someone cut us off on the highway or why we did poorly on an exam, people are always trying to create order out of a chaotic world through the use of causal explanations. As social psychologists observed this behavior more and more they began to formulate a theory to explain how we answer the question “why?” One theory that explains how people make causal attributions is known quite simply as attribution theory (Heider, 1958; Jones & Davis, 1965; Kelley, 1967; Weiner, 1985). This theory has undergone several reformulations and has been incorporated in many scientific studies in many contexts and environments. Attribution theory has been most widely utilized in the classroom and school settings because of its important implications for academic behavior. From children’s unpopularity in school (e.g., Aydin, 1988) to children’s beliefs about the causes of success and failure in academics (e.g., Frieze & Snyder, 1980), attribution theory has been used to help explain the attributions people make in many facets of the educational experience. The current study utilizes the theory to help explain college student’s attributions in a standardized test situation.

The power of standardized testing in today’s American society is profound. Ever since the early 1900’s with the advent of Binet’s Intelligence Quotient and Thorndike’s Handwriting Scale, standardized testing has steadily become a widely used educational measurement tool. The most recent boom in the use of standardized testing came about due to the No Child Left Behind (NCLB) Act of 2001. This legislation calls for, among other things, annual testing of students in reading and math in grades 3 through 8. Even before
NCLB, the uses of standardized tests were extensive. At present, test results have become the basis for selection and retention in numerous educational programs and grade levels. They are used to hold educators and school administrators accountable for their curriculum by appropriating funding based on the scores. And, they have set the standards in almost all subject areas and grade levels, including collegiate and professional training. Because of the widespread nature of standardized tests, there have been many questions asked, criticisms raised, debates argued, and research studies done about how such testing is affecting the education of America’s students.

The current study has set out to provide more information on standardized testing. However, the focus of this study is not on the tests themselves. Rather, this study focuses on one aspect of standardized testing that is relatively overlooked in the literature—the presentation of results to the student. More specifically, the current study is concerned with examining how the way in which results are presented to the student affects their perception of success and failure as well as their future outcome expectations on similar tests. Attribution theory is utilized due to its eclectic nature in that it includes aspects of other widely supported theories. For instance, it incorporates behavior modification in the sense that it emphasizes the idea that students are motivated by the rewarding outcome of being able to have a sense of accomplishment and to feel good about themselves (Weiner, 1982). The theory also integrates cognitive and self-efficacy theories by emphasizing that students’ perceptions of their ability will influence the ways in which they will explain the outcome of their most recent efforts and subsequently their future inclination to perform these same behaviors (Weiner, 1980).
Another facet of this study is the examination of how students attribute their performance based on their health. Numerous studies have looked at how health affects academic performance (e.g., Datar, Sturn, & Magnabosco, 2004; Mo-suwan, Lebel, Puetpaihoon, & Junjana, 1999), but there exists a lack of information on how health-related variables affect attributional style. Of the ones that have addressed the latter question, the most prevalent hypotheses have concerned the relationship between depression and attributional style (e.g., Bennett & Bates, 1995; Gibb & Alloy, 2006). The current study will attempt to fill in this gap in the literature by examining how the physiological health variable body mass index (BMI) affects a student’s performance and their subsequent attributions to that performance.

Each of the above topics, and more, will be discussed in further detail in the following sections.

Health and Academic Performance

A vast number of health-related concerns have been examined to see how they affect academic performance. For example, Alaimo, Olson, Frongillo, and Briefel (2001) found that food insufficiency is a health concern for U.S. preschool and school-aged children. More specifically, the study found that food insufficiency among 6- to 11-year-old children is associated with a .02 and .07 SD decrease in reading and math standardized test scores, respectively. Diabetes, another health concern among children, was found to have a nonsignificant relationship with children’s cognitive skills (McCarthy, Lindgren, Mengeling, Tsulikian, & Engvall, 2003). For most children with Type 1 diabetes, medical conditions associated with diabetes (e.g. blood glucose levels) were not found to be as strongly associated with academic performance as were factors such as socioeconomic status and
academic behavior (e.g. absenteeism). The research findings on asthma yielded similar results. Gutstadt et al. (1989) found that childhood asthma has no significant association with academic performance. While all of these studies yielded relatively low effect sizes, studies on childhood obesity and overweight have found somewhat more significant and consequential findings.

Childhood obesity and overweight have become major health concerns in recent years (National Center for Health Statistics, 1998-1999; U.S. Department of Heath and Human Services, 2001). “According to recent estimates, the prevalence of overweight (BMI ≥ 95th percentile for age and gender) is 15.3% in 6- to 11-year-old children and 10.4% among 2- to 5-year-old children, compared with 11.3% and 7.2%, respectively, in 1988 to 1994” (Datar et al., 2004, p. 58).

Some studies have documented the risk factors associated with obesity. For example, Must and Strauss (1999) found that childhood obesity is related to neurological, pulmonary, gastroenterological, orthopedic, and endocrine conditions. Obesity has also been shown to be related to low self-esteem and depression (Erickson, Robinson, Haydel, & Killen, 2000; French, Story, & Perry, 1995). More relevant to the current study is obesity’s and overweight’s effect on academic performance. On this point, Falkner et al. (2001) found that obese adolescents consider themselves to be worse students than their normal weight peers. This study lacks in some generality, however, in that it is unclear whether the findings hold for younger children or when objective measures of school performance (test scores, class grades, etc.) are used. A second study found that obese children in China had significantly lower IQ scores than a control group (Li, 1995). Another international study that sheds some light on the obesity-academic performance relationship is Mo-suwan et al.
This cross-sectional study found that being overweight in grades 7-9 was related to poorer academic performance, whereas being overweight in grades 3-6 was not related to poorer academic performance. Datar et al. (2004) found potentially conflicting results in their finding that obesity is related to a child’s academics in the younger grades. The study found that overweight kindergarteners scored lower than their non-overweight peers on standardized tests. However, once the study controlled for potential confounders such as mother’s education, race/ethnicity, behavioral factors (e.g. home environment, exercise, and television watching), and baseline achievement scores, overweight status became more of a “marker, but not a causal factor, affecting academic performance” (p. 67). This statement was true for all conditions except boys’ math scores at baseline. However, while this study did conclude that the significant differences in test scores as a result of overweight status could be explained with the other aforementioned variables, overweight status is still more easily observable, especially in younger students.

The literature suggests that obesity is more strongly related to academic performance as students get older. This is potentially due to the effects of socioeconomic status, home environment, and other environmental variables taking further hold and becoming more evident in the later academic grades. The current study reexamines some of the conclusions made by the research by looking at college students. This study also examines how obesity affects the types of attributions students make to their outcome in a testing situation. In order to understand this latter focus better, a discussion of attribution theory is necessary.

The Many Dimensions of Attribution Theory

Attribution theory was developed over time primarily from the theories of social psychologists Heider, Jones, Davis, Kelley, and Weiner. Heider first wrote about attribution...
theory in his book *The Psychology of Interpersonal Relationships* (1958). In this book, he described what was called “naïve” psychology. Heider believed that people’s actions are based on their beliefs. Therefore, these beliefs must be considered if psychologists are to explain human behavior. This would be the case whether the beliefs were true or not. Heider also suggested that people can learn a great deal from what he termed commonsense psychology. This emphasis that Heider placed on the importance of taking the ordinary person’s explanations and understanding of events and behaviors seriously led to other researchers investigating this new line of thought.

*Correspondent inference theory* (Jones & Davis, 1965) was the next step in attribution theory’s formulation. The theory describes how an “alert perceiver” might use another person’s behavior to deduce their intentions and dispositions. The name of the theory comes from people inferring another’s dispositions directly from the other person’s behavior. Inferences are considered correspondent when the behavior and the disposition align and can be assigned similar labels.

*Kelley’s model of attribution theory* (1973) deviates slightly from the previous two in that it is not limited to interpersonal perceptions. Kelley’s model is also unique in that it suggests that people examine three different kinds of information in their efforts to establish the accuracy of their attributions (Ross & Fletcher, 1985). The first type of information is consensus information. This type of information refers to how well a person’s response to a given stimulus matches with the majority of other peoples’ response to the same stimulus. The second type of information is known as distinctiveness information, which refers to the differentiation of responses by the target person to other stimuli. In other words, does the person respond in the same way to other stimuli as well? The third type of information is
consistency information. Broadly, this refers to the degree to which the target person responds in the same way to a given stimulus. By adding these three factors that affect the formation of attributions, Kelley advanced Heider’s original theory.

While Kelley was advancing Heider’s attribution theory through adding three factors that affect the formation of explanations, other researchers and theorists were adding to the theory in another way—causal dimensions. The first causal dimension was originally formulated by Heider (1958) and has already been alluded to. This is most commonly termed the locus of control dimension (Rotter, 1966) and refers to the internal-external nature of attributions. A second dimension is called stability (Weiner et al., 1971). While Heider did allude to this dimension by differentiating between dispositional and relatively fixed characteristics, Weiner et al. (1971) labeled it and incorporated it into the achievement domain. The term refers to how fixed the explanation is. Another dimension of causality that was identified by Heider and later incorporated into the achievement domain by Rosenbaum (1972) is intentionality. However, Weiner (1979) contends that this dimension is mislabeled. He argues that labeling this dimension as controllable would be more appropriate. Rosenbaum labeled it as intentionality because he believed that attributions to mood and effort differ only in their intent, with mood as unintentional and effort as intentional. However, Weiner argued that “failure attributed to a lack of effort does not signify that there was an intent to fail…Rather, effort differs from mood in that only effort is perceived as subject to volitional control” (p. 6). Table 1 organizes these three dimensions and provides some example attributions for each category. Table 1 has been adapted from Weiner (1979).
While these three dimensions (and arguably intentionality) have been identified as the main dimensions used in achievement-related contexts, there is another dimension that has emerged. This dimension is termed *globality* (Abramson, Seligman, & Teasdale, 1978).

This dimension lies on a continuum from global to specific. For example, a student’s ability may be perceived as specific (“I failed this test because I do not understand algebra”) or as general (“I failed because I am dumb”).

Table 1

*Causes of Success and Failure and their Corresponding Dimensions*

<table>
<thead>
<tr>
<th></th>
<th>Internal</th>
<th>Unstable</th>
<th>External</th>
<th>Unstable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controllability</td>
<td>Stable</td>
<td>Unstable</td>
<td>Stable</td>
<td>Unstable</td>
</tr>
<tr>
<td>Uncontrollable</td>
<td>Ability</td>
<td>Mood</td>
<td>Task difficulty</td>
<td>Luck</td>
</tr>
<tr>
<td>Controllable</td>
<td>Typical effort</td>
<td>Immediate effort</td>
<td>Teacher bias</td>
<td>Unusual help from others</td>
</tr>
</tbody>
</table>

The current study narrows these attributions down to ability, immediate effort, task difficulty, and luck. Research has shown that among these four choices ability (stable) and effort (unstable) predominate as the most widely utilized attributions, especially in success conditions (Frieze & Snyder, 1980). This point should be kept in mind when evaluating the hypotheses that will be presented later.

According to attribution theory, the explanations that people make to explain success or failure can be differentiated into four primary categories. First, the cause of success or failure may be internal or external. That is, success or failure may occur due to factors that have their origin within a given person or because of factors that originate in the environment. Second, the cause of the success or failure may either be stable or unstable. If
the cause is deemed to be stable, then the outcome is likely to be the same if the same behavior is performed again. If it is deemed to be unstable, the outcome is likely to be different on another occasion. Third, the cause of the success or failure may be either controllable or uncontrollable. A controllable factor is one which is believed to be alterable. An uncontrollable factor is one that is not believed to be easily altered. Finally, causal explanations can either be global or specific. If the cause is deemed global then the implication is that the same cause will affect other aspects of life. If the cause is deemed specific then it is limited to the task and does not imply any generalities about the person.

The generation of causal attributions is linked to the type of situation a person is in (Anderson, 1983). More specifically, different types of causes are generated for different types of situations, and that these different types of situations lead people to make attributions that differ in dimensionality. With this in mind, because globality is more widely utilized in learned helplessness models (see Abramson et al., 1978) it has not been incorporated into this study. Controllability will also not be utilized because it has not been widely incorporated in achievement contexts. The other two dimensions, however, have most widely been used in attribution models of achievement motivation and expectancy change (Weiner, 1979). In support of this point, Meyer and Koelbl (1982) set out to identify the dimensions underlying students’ causal attributions for their test performance. The study found that Weiner’s conceptions of stability and locus accounted for most of the variance in the participants’ ratings. Because these two causal dimensions have been identified with a relatively high degree of certainty in achievement contexts, they are the dimensions the current study incorporates. In order to better understand how these dimensions play out in
achievement settings, an examination of some research studies that examine just that are necessary.

First, Flammer and Schmid (2003) interviewed children aged 5 to 14 and asked them about causes that lead to success or failure on school tests. After the participants provided open-ended attributions to a success or failure outcome experienced by a character in a fabricated scenario, the attributions were coded and 30 different categories were formed. Of these 30 different categories, practice, specific abilities, and concentration were the most frequently mentioned in both the success and failure conditions. The authors make the point that “if one inserts [these] 30 basic categories into Weiner’s original four-cell table, all the cells are filled” (p. 349) with the internal/stable category containing the largest number of categories. Another point of interest in this study is that the gender differences in attributions were more pronounced when the participant was in the failure condition. In this condition, boys stated that it was the lack of instant effort that counted, whereas girls stated that it was the lack of lasting effort. This study also looked at age differences and found that adolescents mentioned more differentiated categories of causes than their younger counterparts.

Another study looked at 360 children ranging in age from approximately 7-14 years of age in two school systems in India (Sushila, 1988). This study presented children with achievement stories that involved a success or failure condition. Ability- and effort-related causes generated by the participants support Weiner’s model of attribution, whereas task difficulty and luck were rarely used. This study also examined age differences in the formation of causal attributions. While all of the children made ability attributions, older children were more concerned about the reasons for doing well, whereas the younger
children relied more on the outcome. Age differences were not significant when explaining failure. Also, sex differences were not found to be significant in this study.

A third study was performed with 1st-, 3rd-, and 5th-graders who were asked to produce causes for success and failure for four situations: performing on a school test, doing an art project, playing football, and catching frogs (Frieze & Snyder, 1980). The only situation of interest for the current review is the school testing situation. However, it is important to note that causal explanations were found to differ across the four situations. The testing situation was predominated by internal attributions to effort and ability. In fact, only 20% of the causal explanations used by the participants to explain success or failure on a test were not ability or effort. Also, children made the greatest differentiation for success and failure in the testing situation. Age differences were found in that older children made more use of effort attributions and relatively less use of ability attributions, with younger children having fewer codable responses. Also, no significant difference in attribution style between the genders was found.

A fourth study looked at the application of attribution theory to African Americans (Graham, 1988). Among other things the study looked at the differences between Caucasian and African American students in terms of their perceived causes of success and failure on school exams. In terms of failure, both racial groups predominately identified lack of effort as the predominant cause. Similar results occurred in the success condition.

And finally, Yamauchi (1988) studied Japanese undergraduates as they completed various mental tasks in a competitive environment. The study examined the role of the actor and the role of the observer (the opponent) in making causal attributions. In order to study these roles the students were asked to rate the extent to which they attributed their outcomes
and their opponent’s outcome to the four commonly used causes: ability, effort, task
difficulty, and luck. The study also examined the affect of the self-serving bias on the actor’s
role and the observer’s role. According to the notion of self-serving bias, people attribute
success to internal factors and failure to external factors. In this study, the winning actors
attributed success to luck, while the losing actors attributed failure to ability or internal
factors. The results of this study indicated no self-serving bias, suggesting instead a reverse
trend. The losing opponent-observers attributed the actors’ success to more internal factors,
while the winning opponent-observers attributed actors’ failure more to luck.

The preceding studies were selected not only because they illustrate Weiner’s
attributional model at work, but because they examine other variables and concepts that may
have an effect on attribution formation and that could potentially affect the results of the
current study. These will be discussed in further detail in the following section.

The Self-Serving Bias and Other Potential Variables of Interest

Self-Serving Bias

The types of attributions people make have strong implications for self-esteem
(Weiner, 1985). In general, people make attributions that best protect their self-esteem. This
commonsense phenomenon was first introduced by Miller and Ross (1975) as the self-
serving bias. In its simplest terms, the self-serving bias is the tendency to attribute success to
internal facts such as typical effort or ability and failures to external (environmental or
situational) factors such as task difficulty or luck. In this way, the person is able to
internalize successful outcomes while distancing themselves from failures, thereby
preserving their self-esteem. Thus, a common assessment of a person’s own success might
be: “I got an A on this paper because I am a very skilled writer” (ability), or “I did well on
this test because of all of the hours I put into studying” (effort). Common assessments of failure might be: “I didn’t finish the test on time because there was not enough time to answer the questions” (task difficulty), or “I didn’t make it to class on time today because of traffic” (luck/chance).

In their study, Miller and Ross (1975) actually found little support for the idea that the self-serving bias affects causal attributions. However, they did find some support for the argument that people engage in self-enhancing attributions after a success outcome. But only minimal support was found for the use of self-protecting attributions after failure.

In spite of these earlier findings, more recent studies have found a self-serving bias when formulating attributions. One of the most comprehensive among these is a study done by Mezulis, Abramson, Hyde, and Hankin (2004). This study was conducted primarily to address the debate that surrounded the contention that the self-serving bias is widespread among most humans in most cultures. The meta-analytic study addressed the debate by examining 266 published research studies. Some of the results of the meta-analysis indicated that, in general, there were no differences between men and women in their self-serving biases meaning that men and women were just as likely to make self-serving attributions. Additionally, these researchers found that the United States and other Western nations (Canada, the United Kingdom, Australia, New Zealand, and other Western European countries) had a strong self-serving bias, which was more pronounced than in most other cultures on other continents. However, despite these strength-related differences, the researchers found that there was a positive self-serving bias in all cultures studied. Within the United States, there were no meaningful differences in self-serving bias among different racial and ethnic groups; no one race was more likely than the others to be more susceptible
to this self-serving bias. The general conclusion of Mezulis and colleagues was that there is a universal self-serving attributional bias that exists across gender, race, and even nation. The average effect size for all of the studies incorporated in this analysis was 0.96, indicating that a large bias exists.

Kudo and Numazaki (2003) found that participants in a success condition made more internal attributions than their failure condition counterparts. The findings of this study were somewhat surprising because previous research had generally not found a self-serving bias in Japanese participants (see Kurman, 2003). However, once procedures were set in place that reduced evaluation apprehension, the previously mentioned findings emerged.

Another study took an interesting approach to the self-serving question. MacDonald and Standing (2002) asked the question: “Does self-serving bias cancel the Barnum effect?” The Barnum effect refers to people finding personal validation in a statement that could apply to many people. This study found that the self-serving bias was, in fact, powerful enough to cancel out this effect.

After comparing the results of these studies, it is clear that the self-serving bias is a robust phenomenon that needs to be taken into account in the interpretation of results in attributional research.

**Gender**

A second factor that may influence attributions is gender. Gender is one of the most debated factors in attribution research. Research that has examined this topic has come to different conclusions as to whether or not sex differences exist. Literature on sex differences in attributions is inconsistent and has not yet been able to adequately explain differential achievement in men and women (see Frize, Whitley, Hanusa, & McHugh, 1982, for further
discussion). However, in spite of this uncertainty, many studies and researchers have asserted that sex differences do exist and that there are models that explain these differences (see Frieze et al., 1982). McHugh, Frieze, and Hanusa (1982) argue that there are several reasons for the unclear nature of the gender literature. First, many studies imply that one or more of the sex differences models are proven fact even though Frieze et al. (1982) found no substantial support for any of the models. Another reason is that some studies show support for one of the models without mentioning the other possible alternatives. Also, because the models are somewhat ambiguous it is easy for researchers to claim that the models support their findings. Finally, McHugh et al. (1982) argues that publication biases may be present in favor of studies that find sex differences over ones that do not.

What are these models that have been mentioned? There are three basic models of sex differences in attribution research: general externality (Feather, 1969), self-derogation (e.g., Nicholls, 1975), and low expectancy (Deaux, 1976). The general externality model explains female achievement by suggesting that women tend to have an external bias in the attributions they make (Feather, 1969; Simon & Feather, 1973). One possible explanation for this finding is that women withdraw from achievement situations altogether due to their higher proneness to fear of failure and fear of success. Because of this withdrawal it is reasonable to make the supposition that women will tend to view external factors as affecting them more than internal ones. This general external attributional style protects women from fears of success by removing responsibility and limiting feelings of shame for failure (Simon & Feather, 1973).

Self-derogation is a second model used to explain sex differences. According to Nicholls (1975), women tend to attribute their success to external factors whereas they
attribute their failure to internal factors. According to Aronson and Mettee (1968), this is because people desire consistency in their self-beliefs. In other words, if someone has low self-esteem, they are willing to believe on negative information about themselves, and vice versa for high self-esteem. Adding to this, Frieze et al. (1978) states that women typically have relatively lower self-esteem in achievement settings as compared to men. Therefore, it can be reasonably concluded that women will tend to incorporate more negative information into their self-belief system (internal) in achievement situations than their male counterparts.

The low expectancy model posits the idea that women generally have lower expectations about their performance on achievement tasks and in achievement situations (Deaux, 1976). These low expectations lead to unstable attributions for success and stable attributions for failure. Then, as a result of success outcomes being discounted, there are no increased expectations for future tasks and the low expectations are therefore reinforced and perpetuated.

All of these models were examined by Frieze et al. (1982) in a meta-analytic review of twenty-one studies. They found that none of the aforementioned models was well supported by the research. In terms of the failure to reach a reasonably consistent finding on sex differences in individual studies, there have been two explanations offered: “the failure to adequately consider various situational determinants of sex differences in attribution, and the problems of viewing women as an homogeneous group” (McHugh, Frieze, & Hanusa, 1982, p. 469).

A final word on this matter comes from Sohn (1982), who conducted a meta-analytic effect-size analysis on findings from studies that examined sex differences in achievement attributions. It was found that, with the exception of luck attributions for success, all other
effect sizes of gender on attributions were less than .01. This finding suggests that even if a study has found statistically significant sex differences, these differences may not be consequentially significant.

**Race/Ethnicity**

Studies done on the effects race has on attribution making have primarily focused on differences between European Americans and African Americans. However, one study that did not focus on these two groups examined on how people of different ethnic groups and socioeconomic statuses would judge an actor’s behavior—either socially desirable or socially undesirable (Mann & Taylor, 1974). The study incorporated middle-class English Canadians and French Canadians. These participants were asked to rate the relative importance of the internal traits of actors in causing them to behave in certain ways after the actor was described as being French or English Canadian and low- or middle-class. An ingroup bias was found in that participants made more favorable attributions to members of their own ethnicity and class than for members of the outgroups.

Phinney, Baumann, and Blanton (2001) also did not solely focus on African Americans and European Americans. This study also went a step further in that it not only examined the effect ethnicity has on attributions, but also its effect on life goals and outcome expectancy. The sample consisted of 371 adolescents ranging in age from 12-18 years of age. The adolescents came from five different ethnic backgrounds: Armenian, Vietnamese, Mexican American, European American, and African American. The participants responded to open-ended questions regarding their future goals, their expectations for reaching those goals, and their attributions for the expected outcome. Results show that, in general, adolescents set high goals and expected to reach them. Of those who expected to reach their
goals attributed the outcome more to effort and less to external factors than did those who did not expect success, as would be suggested by the self-serving bias. Compared to European American adolescents, minority group adolescents attributed outcomes more to effort and less to ability. The authors of this study postulated that this last finding may reflect the belief of minority group adolescents need to work harder to overcome obstacles caused by stereotypes.

However, as noted before, the majority of the studies done in this domain have focused on European and African Americans and their differences with respect to perceived causes of success and failure. An early set of studies found that African American children less frequently use effort as an attribution in situations with rewards and punishments (Weiner & Peter, 1973; McMillan, 1980). Graham (1988) notes that these findings were “compatible with the then dominant view in the motivation literature of blacks as more externally oriented than whites” (p. 7). Katz’s (1969) ideas can be used to interpret this result with the supposition that African Americans place less value on effort as an indicator of performance. Another early predominant finding in the research suggested that African Americans possess a more adaptive attributional style (Graham, 1984). This means that they attribute failure to lack of effort, perceive themselves as being more competent, and persist longer at a task. Graham’s study found this attributional pattern among middle-class African Americans following a failure outcome on a puzzle-solving task. However, more recent studies have not found these differences (see Graham, 1991 for a comprehensive review). The more recent findings have shown no racial differences in attributional style between African and European Americans and no differences favoring African Americans over European Americans (Graham, 1984). While Graham’s findings may seem contradictory,
the adaptive attributional style was only found after controlling for affective cues and was only found in middle-class blacks. In general, however, no significant racial differences were found.

In her review, Graham (1988) makes the statement that:

“...it often is the case that when we find differences between blacks and whites – say, for example, less of an achievement motive in blacks and less capacity to delay gratification – we dismiss them as irrelevant, myopic, ethnocentric, or outright racist. On the other hand, when we find similarities between the two racial groups, then often the case is closed” (p. 3).

Graham is referring to the use of a comparative approach and how this approach has not led to any sufficient progress in terms of gaining knowledge about motivation in blacks due to the societal bias to find similarities. Because of this concern, the question about racial differences in attribution making may still not be as conclusive as the more recent studies on the subject may suggest. However, in spite of this concern and in light of the relatively recent findings that there are no racial differences in attribution making, the current study has not included race/ethnicity in its analyses. The other reason for this (which will be discussed again in the method section) is that the population for the current study is relatively homogeneous and therefore there is not a large enough racial disparity to sample from in order to yield any significant findings.

Age

Studies that have examined age differences in achievement attributions have yielded more consistent findings as compared to gender. While a meta-analytic review of this topic was unable to be identified, many individual research findings exist. These will serve as the basis of the discussion in this section.
Sushila (1988) found a variety of differences in attributions based on age. This study found age differences in the use of performance ability, specific competence, general competence, effort, interest, others, and behavior as causal attribution categories. More specifically, younger children tended to use performance ability and others more frequently than the older children. In general, “the younger was the child, the stronger was the tendency in the subjects to focus on concrete features and adopt more global categories” (p. 10). The study also found that age differences were only significant in explanations of success and not failure. The author notes how some of the findings fit in with a Piagetian framework. The move from general to specific attributions and the increased use of effort (as a category) with age fit well with Piaget’s (1932) ideas that as children age they gain an enhanced awareness of specific skills and that the use of effort becomes more important in school achievement situations, especially with rewards.

Another study tested the relationship between age and self-other differences in attribution formation (Skinner, Schindler, & Tschechne, 1990). They conducted a series of studies to identify when and in what form in a child’s development would self-other differences appear. It was found that there existed no self-other differences in how children attributed school performance before the age of 10. Starting at age 10 to 11 (grade 4), however, self-other differences began to emerge. This emergence primarily took form in the internal causes, such as effort and ability. No mean level differences were found for external causes. In addition, perceived control was only found in beliefs about the self. A final important finding of this study was that self-other differences have a strong relationship with academic performance, primarily for beliefs about the self.
A third study conducted by Clayton et al. (1992) deviates from these previous studies in that it examined the antecedents of making attributions and the differential effect of age. This cross-sectional study looked at students from year 4, 6, 7, 9, 11, as well as TAFE (Technical and Further Educations) classes. Each of the classes was given a scale that included three of the classic four attribution categories (ability, luck, effort) as well as a fourth category termed strategy attributions (for a further discussion of strategy attributions see Chan & Moore, 2006). The students were asked to rate these categories in success and failure situations. The findings were that for students in grades 4 and 6, attributions to effort in success situations were positively related to achievement in math and English. For students in year 9, however, the positive predictor of achievement was found to be strategy attributions. A final result of this study was the finding that ability attributions for failure were always a negative influence on achievement, except for students in year 4.

Other interesting results with regard to age differences include the finding that as students get older, attributions to bad luck and other external factors decline (Powers & Wagner, 1984). Also, O’Sullivan and Howe (1996) found that attributions become increasingly better predictors of achievement as children grow up.

The aforementioned studies not only shed light on how students of different grade levels use attributional categories to explain success and failure, but also what affects these attributions have on the students with regard to their level of achievement. The general findings from these studies are that as students grow older they become more differentiated and complex in their attributional style, and that this greater intricacy leads to attributions becoming more strongly tied to achievement.
Attribution-Expectancy Link

Attribution theory is a commonly used theory to explain the dynamics of motivation, it is also inextricably tied in with expectations (Weiner, 1985). There have been many motivationally based theories as to how this link functions. Heider (1958), for starters, stated that expectancies in achievement related contexts are determined by perceived ability and the amount of effort put forward, relative to the perceived difficulty of the task. Other notions on this topic have looked at expectancy as a function of reinforcements (e.g., Rotter, 1966). However, Atkinson (1964) postulated a slightly more comprehensive conception. He saw expectancy as influenced by the number of individuals against whom one is competing, prior reinforcement history, and communications from others concerning the likelihood of success.

An attributional approach to expectancy change has yielded more conclusive findings. In fact, Weiner (1985) developed an expectancy principle based on causal attributions and three corollaries. This principle is based on the conceptual link between the stability dimension and expectancy. It was contended at one point that locus of causality was the better determinant for expectancy change (Feather, 1969; Rotter, 1966), which is still considered partially true, but it was later found that the stability of a cause better determined expectancies of success and failure (see Weiner, 1985). Weiner’s (1985) principle and corollaries underscore this relationship.

*Expectancy Principle*: Changes in expectancy of success following an outcome are influenced by the perceived stability of the cause of the event.

*Corollary 1*: If the outcome of an event is ascribed to a stable cause, then that outcome will be anticipated with increased certainty, or with an increased expectancy, in the future.
Corollary 2: If the outcome of an event is ascribed to an unstable cause, then the certainty or expectancy of that outcome may be unchanged or the future may be anticipated to be different from the past.

Corollary 3: Outcomes ascribed to stable causes will be anticipated to be repeated in the future with a great degree of certainty than are outcomes ascribed to unstable causes.

Numerous research findings have supported these postulates. For instance, Kovenklioglu and Greenhaus (1978) studied the link between causal attributions and expectancies in college students. Participants took two chemistry tests at two different time points and their outcome was classified as a success or failure. The students then attributed their performance outcome to the classic four causes: ability, effort, luck, and task difficulty. Following a successful outcome, students’ expected and actual performance were positively related to attributions to high ability (stable) and negatively related to good luck (unstable). On the other hand, among students who failed the tests, expected performance was positively related to attributions to low effort (unstable) and negatively related to attributions to low ability. Remember, there are two types of effort that were distinguished in Table 1. Typical effort was considered to be stable while immediate effort was considered to be unstable. In this study, immediate effort was the type of effort under investigation. Another significant finding of this study was that expectations were found to be strongly related to performance on the second test. However, the authors do note that this finding was diminished, but still significant, when attributions to ability and past performance were held constant.

A second study looked at expectancy change after a confirming or disconfirming outcome (McMahan, 1973). The author of this study presented five anagrams to 6th-graders,
10th-graders, and college students. The participants rated their expectancy of success prior to each anagram and their causal attributions for the success or failure afterwards. It was found that expectancy disconfirmation led to more attributions to effort and luck and fewer attributions to ability. Also, attributions to ability and task were associated with high expectancies following success and with low expectancies following failure. This is in contrast to attributions being made to effort and luck under conditions of low expectancies following success and with high expectancies following failure. A general finding of this study is that students tend to give unstable causes following expectancy disconfirmation, whereas the tendency is to give stable causes following expectancy confirmation.

These findings are in line with the general concept of the relationship between causal attributions and expectancy of success (or failure). Attribution theory proposes that the level of expected performance depends on the causes attributed to the initial success or failure. It has been shown that stability is the causal dimension most related to expectancy. The rationale is that stable factors are perceived to continue into the future, thus making future success or failure more certain. From a motivational perspective, it makes sense to see failure as having unstable sources and success as having stable sources. If we believe that our success is due to something that is not necessarily going to happen again then we will have as high of expectancies for future performance and will only motivate ourselves enough to do what is necessary to reach this relatively lower expectancy level. Another factor in this attribution, expectation, and motivation relationship is discussed in the next section.

Task- and Ego-Involving Situations

In the aforementioned study conducted by Skinner, Schindler, and Tschechne (1990), the issue of self-other differences in attribution was broached. While the current study does
not ask its participants to make attributions about other students, it does contain two conditions: one that compares participants’ test scores with that of other students (norm-referenced) and one in which the participant’s score is presented by itself; without reference to other scores (criterion-referenced). In this sense, a similar vein of research to self-other differences is addressed. This line of research, first introduced by Nicholls (1975) differentiates between situations that are ego-involving and situations that are task-involving.

Task-involving situations are those where one’s goal is to master the task. Greater understanding or acquisition of new skills is considered an end in itself. For example, a situation where a student receives personalized feedback, praise, and/or is rewarded for high levels of effort is task-involving. This is contrasted with ego-involving situations where one’s primary goal is to demonstrate high ability relative to others in the case of perceived success or to conceal low ability in the case of perceived failure. For instance, a situation where a teacher gives most of her or his attention to the students with higher ability or where students’ performance is directly compared to their peers are considered ego-involving situations.

Research that has studied this distinction has found that task-involving situations can be antecedents to high ability, whereas ego-involving situations can be antecedents to low ability (Jagacinski & Nicholls, 1984). For example, in a task-involving situation, if one succeeds after giving a high level of effort, then this leads to the attribution that one is also high in ability. In other words, success following from high effort suggests that one has acquired new skills and therefore has increased personal competence. Moreover, attributions to effort are most prevalent after a success outcome in task-involving situations. Research has shown that in task-involving situations, success is derived from trying hard as well as
working well with others (Ames & Ames, 1984; Seifriz, Duda, & Chi, 1992). With ego-involving situations, on the other hand, high effort followed by success is more often a cue to low ability, particularly when individuals compare their performance to that of others. Therefore, attributions to ability are most common in ego-involving situations (Jagacinski & Nicholls, 1984). The findings that task-involving situations typically lead to attributions to effort and ego-involving situations typically lead to attributions to ability in success situations is a second important point to note when evaluating the hypotheses. This point is coupled with Frieze and Snyder’s (1980) finding that ability and effort attributions are the foremost attributions made in achievement situations.

Another relevant aspect of the literature comes from Butler (1987). The author of this study suggested that written teacher comments to students about their performance on an assignment encouraged task involvement, whereas the use of letter grades about the same performance promoted ego involvement. In support of this hypothesis, the study found that written teacher comments led to more interest in the task, better performance, and higher attributions to effort. This was in contrast to normative grades, which resulted in less interest, low performance, and more attributions to perceived ability, or lack thereof. In this study, the letter grades were considered normative information because they were seen as encouraging social comparison in the classroom. While some of the participants in the current study may have discussed their score with other participants and therefore became more ego-involved in the situation, this certainly was not the norm due to the noncompetitive atmosphere.

In line with these results, but one step further, are findings that suggest that internal attributions of achievement are associated with higher actual achievement (e.g., Georgiou,
The former study found that underachievers tended to attribute their academic performance to external factors, while higher achieving students tended to attribute their performance to internal factors. Similar findings were found in the latter study where 3rd-, 6th- and 9th-graders were asked to rate the importance of several causal variables for their performance on a reading test. The higher the student’s reading achievement, the more attributions were made to internal causes.

One of the implications of these findings is that individuals who believe that their own efforts are ineffective compared with those of other people may become passive and helpless, especially if they attribute the outcome to a lack of ability (Abramson et al., 1978). In other words, students are likely to discount their competence if met with competitive evaluation grading systems. Therefore, if students are to persist at academic tasks, they should be encouraged to establish a genuine belief that they are competent and that imperfections or failures are the result of some other (external) factor. One practical way to do this is in the interpretation of test results to students. The next section will address this topic.

Criterion- and Norm-Referenced Feedback

As students grow older and graduate through grade levels, test results become more useful to the student especially for self-evaluation purposes (University of Iowa College of Education, 2007). The scores on standardized achievement tests can help students decide how much they have learned as well as their strengths and weaknesses in certain general competency domains. Results from standardized tests can be one useful tool to help students in their academic decision-making. These effects are particularly evidenced by the third
grade when students become able to at least rudimentarily understand their test scores (University of Iowa College of Education, 2007).

There are two main types of tests and reporting: criterion-referenced and norm-referenced. According to Popham (1975), these two types differ on four dimensions: purpose, content, item characteristics, and score interpretation. The table on the next page is adapted from Popham (1975) and organizes the primary differences of these two types of tests on the four aforementioned dimensions.

As evidenced in the score interpretation row of the table, the ways in which scores can be presented to students varies. The first methods that will be discussed fall under the norm-referenced category. The first norm-referenced reporting style is known as standard nine (stanine) scores (Eissenberg & Rudner, 1988). These scores range from a low score of 1 to a high score of 9. Scores of 1, 2, or 3 are considered below average, 4, 5, or 6 are considered average, and 7, 8, or 9 are above average. A below average scores means that the child needs improvement in that subject area. This is in contrast to a child who receives a stanine score that is above average. This indicates that the student performed better in that area than other students who took the test.

Percentile scores are a very common way test results are reported. In contrast to stanines, percentiles give the student a more detailed description of how they compare with other students who took the test by showing scores that range from 1 to 99 (Eissenberg & Rudner, 1988). For example, if a student scores in the 80th percentile on a test, that student has achieved a score that is higher than 80% of the other students who took the test. So, if 1,000 students took the test, the student in the 80th percentile scored higher than 800 students.
Table 2

*Criterion- and Norm-Referenced Test Differences*

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Criterion-Referenced Tests</th>
<th>Norm-Referenced Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>To determine whether each student has achieved specific skills or concepts.</td>
<td>To rank each student with respect to the achievement of others in broad areas of knowledge.</td>
</tr>
<tr>
<td></td>
<td>To find out how much students know before instruction begins and after it has finished.</td>
<td>To discriminate between high and low achievers.</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td>Measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts.</td>
<td>Measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts.</td>
</tr>
<tr>
<td></td>
<td>Each skill is tested by at least four items in order to obtain an adequate sample of student performance and to minimize the effect of guessing.</td>
<td>Each skill is usually tested by less than four items.</td>
</tr>
<tr>
<td><strong>Item Characteristics</strong></td>
<td>The items which test any given skill are similar in difficulty.</td>
<td>Items vary in difficulty.</td>
</tr>
<tr>
<td></td>
<td>Each individual is compared with a preset standard for acceptable achievement. The performance of other examinees is irrelevant.</td>
<td>Each individual is compared with other students and assigned a score--usually expressed as a percentile, a grade equivalent score, or a stanine.</td>
</tr>
<tr>
<td><strong>Score Interpretation</strong></td>
<td>A student’s score is usually expressed as a percentage.</td>
<td>Student achievement is reported for broad skill areas, although some norm-referenced tests do report student achievement for individual skills.</td>
</tr>
<tr>
<td></td>
<td>Student achievement is reported for individual skills.</td>
<td></td>
</tr>
</tbody>
</table>
Grade-equivalent (GE) scores are a third method of norm-referenced reporting (Eissenberg & Rudner, 1988). The procedure for obtaining GE scores involves giving a test that is targeted to the average expected ability of a specific grade to students in other grade levels. For example, GEs can be established for a 5\textsuperscript{th} grade test by giving that same test to students who are in the 3\textsuperscript{rd} and the 7\textsuperscript{th} grades. If 5\textsuperscript{th}-grader scores at the 3\textsuperscript{rd}-grade level, then it means that the student reads 5\textsuperscript{th}-grade material as well as a 3\textsuperscript{rd}-grader.

A final method of norm-referenced reporting is ranking (Eissenberg & Rudner, 1988). This simply refers to reporting how well a test-taker performed by giving them their relative position with respect to the other test-takers or a normed reference group, but with no mention of their actual score. For instance, a test-taker may be ranked third out of forty-three.

These three norm-referenced reporting methods all compare students’ scores to the scores of another, large group of students who took the same test (Eissenberg & Rudner, 1988). This comparison group can range in size from other students in the same district to a national sample of students (Eissenberg & Rudner, 1988). Norm-referenced test results are most meaningful when discussed in relation to these other students’ scores.

Criterion-referenced tests, on the other hand, do not require this comparison group to make more objective inferences about a student’s performance. A test that uses a raw score falls under this category. A raw score is tabulated by adding up the number of questions a student gets right. This meaning of this score depends on how many and how hard the questions are. For example, if a student gets 15 right on both a reading test and social studies test, it would not be sensible to conclude that the student’s performance in these two areas is
similar. In order to make the score more readily understandable, raw scores are often converted into other types of scores (i.e. percent correct).

One type of score they are converted into is the percent correct score. Percent correct scores are tabulated by taking the raw score, dividing it by the total number of questions on the test, and then multiplying that fraction (or decimal) by 100. Again, however, without knowledge about the difficulty of the test, percent correct scores are not very beneficial.

Because of the nature of criterion-referenced reporting, it fits well with the concept of task-involving situations. In a task-involving situation, individuals use more self-referenced criteria to judge their success (Standage, Duda, & Pensgaard, 2005). Criterion-referenced reporting provides this type of information and therefore allows for task-involving attributions, most notably to effort. Norm-referenced tasks, on the other hand, place more emphasis on relative ability and this is even marked by differential treatment for people of high and low ability (Newton, Duda, & Yin, 2000).

In short, criterion-referenced reporting involves comparing a student’s score with an objective standard of performance rather than with the performance of a comparison group. The percent correct score is the most commonly used criterion-referenced reporting method while the percentile score is the most commonly used norm-referenced reporting method. Standardized tests that most students take on an annual basis are designed mainly to provide norm-referenced interpretations of the scores obtained from them.

Overview of the Literature and Hypotheses

A unique feature of humans is an aspiration to understand and explain what is going on in the world. This desire means that people can be easily influenced if their explanations for an outcome can be altered. If how people understand and explain what is going on can be
influenced, their behaviors may be influenced as well. This concept is important in the educational setting, especially with younger students. In the younger grades students are very impressionable and a lot of factors can easily influence their academic performance. One predominant factor is standardized testing. With the recent jump in the amount of standardized testing students are exposed to it is important to take a step back and study the effects that such testing has on students. The current study tackles this issue by examining the effects that the presentation of results has on college students’ causal attributions for the outcome and their subsequent expectancy change concerning future performance on similar such tests. College students are being utilized due to the unavailability of elementary-aged students. This is not a limitation in that college students can help shine some light on the issue of standardized testing as they have gone through elementary school and have had a great deal of experience with such testing, most notably with the SAT.

The current study utilizes attribution theory because it has been shown to be related to other constructs that have implications in the academic domain, namely expectancy change (see Weiner, 1985). While there have been many reformulations of this theory, Weiner’s (1979) primary four causal attributions – ability, effort, task difficulty, and luck – will be utilized as these are the most widely tested and supported causal categories in achievement-related situations. The most predominant among these are ability and effort (Frieze & Snyder, 1980). These four causes are delineated on the dimensions of locus, stability, and controllability. For example, an attribution to ability constitutes an internal, uncontrollable, and stable explanation for an outcome, whereas an attribution to luck is external, uncontrollable, and unstable in nature. For the purposes of this study, however, controllability will not be a factor of investigation.
The types of explanations that people make to explain a given outcome are affected by the self-serving bias and age. Other variables of interest that have not shown to affect attributions are gender and race. The self-serving bias refers to people’s tendency to make attributions that best protect their self-esteem and academic self-concept. For instance, a failure outcome would lead to an external explanation while a success outcome would more likely lead to an internal attribution. The research has gone back and forth on whether or not the self-serving bias actually affects the types of attributions people make. However, it has been more conclusively shown that the self-serving bias is a real phenomenon in the attributional domain (Mezulis et al., 2004; Kudo & Numazaki, 2003). The research that has been done on gender and its affects on attribution formation has not yielded conclusive findings (Frieze et al., 1982; McHugh, Frieze, & Hanusa, 1982). Sohn (1982) ran a meta-analytic effect-size analysis on the findings of studies that examined sex differences in achievement attributions and found that most of the effects of gender on attributions had an effect size of less than .01. Race is another variable with the potential to affect attributions. However, just as with gender, the more recent studies on race have not found significant differences (see Graham, 1991 for a comprehensive review). The more recent findings have also shown no racial differences in attributional style between African and European Americans (Graham, 1984), which are the predominately studied races with regard to achievement-related attributions. Age is the final potentially influential variable. In contrast to gender and race, studies examining the effects of age have yielded significant findings (e.g., Clayton & Jones, 1992; O’ Sullivan & Howe, 1996; Powers & Wagner, 1984; Skinner, Schindler, & Tschechne, 1990). The general finding in these studies is that as students grow older attributions become more complex and more strongly tied to achievement.
The type of situation a person is in can also affect the causal explanations. Nicholls (1975) differentiates between two types of situations: those that are ego-involving and those that are task-involving. Task-involving situations are those where one’s goal is to master the task. This is contrasted with ego-involving situations where one’s primary goal is to demonstrate high ability relative to others in the case of perceived success or to conceal low ability in the case of perceived failure. Also, following a successful outcome, task-involving situations elicit more attributions to effort whereas ego-involving situations elicit more attributions to ability (Butler, 1987; Jagacinski & Nicholls, 1984).

These two types of situations map well onto the two methods in which scores are reported to students: criterion- and norm-referenced. While there are a number of ways to present scores to students (e.g. percentile scores, raw scores, grade equivalent scores), there are primarily these two broader categories. Criterion-referenced tests compare the students’ scores to a set standard. Percent correct and raw scores are two examples of criterion-referencing. Tests that utilize this method can be characterized as task-involving tests because they focus the standard for success on mastery of the task and allow for self-referenced feedback. This is in contrast to norm-referenced tests, which compare the students’ scores to a standard set by another subset of students. The most common way norm-referenced test results are reported is percentile scores. Because this style of test emphasizes success that is conditional upon the performance of others, it could be characterized as an ego-involving test.

The attribution-expectancy link is of paramount interest in the current study. Different conceptions of what converges to influence expectancies have been postulated and tested (e.g. Atkinson, 1964; Heider, 1958; Rotter, 1966). The most relevant conception of
expectancy relative to this study comes from Weiner (1985). His concept focuses on how attributions affect expectations. More specifically, the focus is on how the causal dimension of stability predicts expectancies of future success or failure. The general idea of his concept of the attribution-expectancy link is that the more stable a cause is perceived to be the more likely that cause will be expected in the future. This is not to overlook the locus of causality dimension and its role, although less significant, in determining expectancies. Using this causal dimension, expectancies will increase, for instance, following a success if the success is attributed to internal factors. If these two dimensions are combined, it can be inferred that attributions that are internal and stable (e.g. ability) will be better predictors of higher expectancy for future success than an attribution to an internal and unstable factor (e.g. effort).

One emphasis of the current study is how the health variable body mass index affects attributional processes as well as performance on the standardized test questions. Previous research has shown that obese and overweight students have lower levels of academic achievement and perceive themselves as worse students (e.g., Falkner et al., 2001). Also, the literature suggests that obesity has a greater affect on academic performance as students get older (Mo-suwan et al., 1999). However, there exists a dearth of information regarding overweight and obesity’s effects on attributions. Therefore, this aspect of the study will be investigational in nature. The current study primarily investigates weight status’s impact on performance. The following hypothesis addresses this focus:

- It is predicted that BMI will be inversely related to performance on the standardized test questions.
The current study also reexamines some of the conclusions made by the research by looking at college students. In particular, the study utilizes Weiner’s (1985) expectancy principle and corollaries to examine the differential effects of criterion- and norm-referenced feedback. With regard to the focus of this study on the relationship between expectancy, criterion- and norm-referencing, and attributions, the hypotheses are as follows:

- Following a success experience, criterion-referenced results are predicted to cause more attributions to unstable causes (namely effort) and, therefore, future expectancy will remain unchanged. Norm-referenced tests are predicted to cause attributions to stable causes (namely ability) and therefore mediate higher expectancy in the future.

- Following failure, it is predicted that there will be more attributions to external attributions which will cause expected future performance to remain unchanged. This is predicted to be the case when presented with either type of feedback due to the pervasive effects of the self-serving bias.

Other explorations of interest are how attributions (particularly to stable causes) and future expectancy are affected by the outcome, how the report type affects attributions to the four causes when not grouped into the stability dimension, and how BMI is related to attributions to the four causes. These investigations serve two purposes: to replicate previous findings as well as to explore other possible effects of the reporting method and health without specific predictions.
Method

Participants

Originally, the participants in this study were going to be 4th- and 5th- grade students from a local elementary school. However, due to unforeseen circumstances, a more convenient sample became necessary. Therefore, the participants for this study are 56 college students from a mid-sized Midwestern university. Students were recruited from introductory psychology courses via an online participant pool called Experimetrix.com. Students in these courses are encouraged to complete twelve credit hours of research as a participant. While they do have other options to get these hours (e.g. writing papers), students typically choose the research route. Because the recruiting was done through an introductory college course the average age of the participants is 19.48 (SD = 1.44). Of the participants, the majority are female (63%). The majority of the participants are Caucasian (84.5%) with the next largest ethnic group being African Americans (5.2%). The average Grade Point Average (GPA) is 3.08 (SD = .57).

Design and Materials

The current study utilized a two time-point between-subjects design. It was not a repeated-measures or within-subjects design because the first time-point did not include any groupings into levels of an independent variable; it only established baseline expectancy and attribution scores, demographics, and test responses. Also, two time points were only necessary to grade the test responses in order to provide participants with their actual scores. The study included one independent variable with two levels. Participants were given their scores in either a criterion- or norm-referenced manner. Students receiving criterion-referenced scores saw their overall score as a raw score and as a percent correct (see
Appendix A). For instance, if a student got 38 out of the 41 test questions correct, they would see their score as all of the above: “38/41,” “93%.” Students receiving their scores in a norm-referenced manner saw their scores as a percentile and as a rank (see Appendix B). For example, a participant may have gotten a score of “90th percentile” and “2nd,” depending on the distribution of scores among the sample. Participants were randomly placed into these two groups with roughly 50% (24-25) in each group.

The items in this study were generated by the primary researcher as well as adapted from other sources. The paired-comparisons attribution question format was adopted from Kovenklioglu and Greenhaus (1978). Participants were given a brief hypothetical success and failure situation and then asked to respond to six forced-choice responses for each situation, for a total of twelve responses that formed the baseline attribution style. An example of this type of item was: “You were NOT successful on the test mainly because.” Then the options were two of the following: you were unlucky, the test was difficult, you did not study enough for the test, and you did not understand the material. All possible unique combinations (for a total of six) were generated for these four response options. Each attribution was scored based on the number of times it was checked (ranging from 0 to 3). This paired-comparisons approach was also previously utilized by McMahan (1973). Reports of the validity and reliability of this method were unable to be found. However, the purpose of using this method was to simplify the task when 4th- and 5th-graders were the population of interest. Due to time constraints and concerns about fatigue effects, the questionnaire was not reformatted after the change in population.

The questions concerning participants’ expectations were generated for this study. An example of this type of question is, “How well do you expect to do on the upcoming math
and reading standardized test questions?” Participants could respond on a five-point Likert Scale from “Poor” to “Excellent.” The SAT questions were adopted from practice exam books (see Appendix C). Participants were given math and reading/writing questions. The math section consisted of both multiple-choice and open-ended response items. The reading section contained sentence-completion questions, whereby a sentence contained blanks and the respondent was asked to fill in the blanks with the most appropriate words, and improving sentences questions, whereby part or whole of a sentence was in parentheses and the respondent was asked to provide the answer that contained the best grammatical choice.

**Procedure**

As already noted, the study took place over two time points. The first time point was conducted in an on-campus computer lab where students were directed to the online survey. Participants had a fixed amount of time to complete the survey (fifty minutes). The first set of questions asked for demographic information, height, and weight. The participants then reached a page that asked them to stop and wait for further instruction. Once all of the participants indicated that they had completed the initial questions, they were prompted to start the standardized test section. Participants were given 40 minutes to complete this section, which allowed a little under a minute per question. The responses to the SAT questions were manually graded and the scores were tabulated and recorded into an Excel spreadsheet.

At the second time point participants received their scores in either a criterion- or norm-referenced format. Those students whose performance on the test was higher than their previously stated minimum standard for success (as given at the first time point) were instructed to describe why they were “successful” by responding to six pairs of success
Attribution items similar in nature to the ones experienced at the first time point (just not hypothetical) (see Appendix D for the follow-up questionnaire that includes these items). These response options were: the test was difficult, I was unlucky, I did not try hard enough on each question, and I am not good at these kinds of tests. Scores for each success attribution were, again, the number of times it was checked. Participants were also asked to rate how well they would expect to do on another similar test in the future.

The self-reported method of assignment to success and failure groups was utilized as it affords greater ecological validity. It was also utilized so that deception would not be necessary. Finally, self-assigned success and failure groups were used so that the participants would not be confused about why they were “successful” or “not successful.” This is well illustrated in a situation where a participant is confident they did well on the test but were then told they were not successful. This would lead to confusion over the success criteria and lead to a potential confound.

The study was approved by the university’s Institutional Review Board. No deception was utilized and all participants gave informed consent before taking part in any aspect of the study and were debriefed upon completion of the second time point.

Results

Before presenting results immediately relevant to the hypotheses, other, more general, findings need to be discussed. Data on the SAT questions are first among these. The average score on the SAT questions were 22 (SD = 7.43) out of 41. Moreover, the average verbal score was 11.21 (SD = 3.85) out of 17, and the average math score was 10.76 (SD = 4.11) out of 24. Table 3 presents correlations among the verbal, math, and total scores.
Table 3

*Bivariate Correlations among SAT Scores*

<table>
<thead>
<tr>
<th></th>
<th>2.</th>
<th>3.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Math Scores</td>
<td>.739**</td>
<td>.937**</td>
</tr>
<tr>
<td>2. Verbal Scores</td>
<td></td>
<td>.928**</td>
</tr>
<tr>
<td>3. Total Scores</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Significant at the .01 level

Another finding of note is how much effort the participants put into answering the questions during the first time point. Participants could rate their level of effort on a 5-point Likert scale ranging from “A great deal of effort” to “Very little or no effort.” Most (53.7%) of the participants reported “Some effort” with 94.4% reporting at least this level of effort. Moreover, effort was positively related to performance, \( r(52) = -0.46, p < .01 \). In reading this correlation it is important to note how the item was coded, with “Very little or no effort” being a 5. Participants were also asked how often they told the truth and answered on a 5-point Likert scale ranging from “Always” to “Never.” Most (72.4%) of the participants reported that they “Always” told the truth, with another 25.9% reporting that they “Almost always” told the truth.

The remainder of this section will be divided into four subsections: three for the specific hypotheses and one for omnibus investigations.

*BMI and Test Performance*

BMI was not found to have an inverse relationship with test performance, \( r(50) = 0.01, p > .05 \). Other findings with regard to BMI are that the average height of participants was 67.70 inches (\( SD = 4.02 \)), and the average weight was 148.57 (\( SD = 28.69 \)).
resulted in an average BMI of 22.85 ($SD = 2.96$), which is within the normal BMI range of 18.5-24.9. Figure 1 breaks down the percentages of participants that fell within the four weight categories. Most (79%) of the participants fell within the normal BMI range. The lack of variability in BMI scores may have influenced the nonsignificant finding on this hypothesis.

![Figure 1. Breakdown by Weight Category (BMI)](image)

**Attributions and Future Expectancy Following Success**

The percentage of participants that attributed their outcome to stable or unstable causes did not differ by report type, $\chi^2(1, N = 49) = 0.34, p > .05$. Specifically, in cases of success, the report type was not found to affect attributions to stable or unstable causes, $\chi^2(1, N = 28) = 0.05, p > .05$. More specifically, there was not a significant effect of the feedback method on attributions to ability, $t(26) = 1.20, p > .05$, task difficulty, $t(26) = -0.60, p > .05$, effort, $t(26) = -0.63, p > .05$, or luck, $t(26) = 0.43, p > .05$, in cases of success. In other words, criterion-referenced feedback did not lead to more attributions to unstable causes (e.g. effort) and norm-referenced feedback did not lead to more attributions to stable causes (e.g. ability).
In spite of these findings, the effect that attributions to stable and unstable causes have on future expectancy is still important to address. On this point, there was not a significant effect of stability on future expectancy, $t(47) = 0.12, p > .05$.

**Attributions and Future Expectancy Following Failure**

The percentage of participants that attributed their outcome to internal or external causes did not differ by a success or failure outcome, $\chi^2(1, N = 49) = 1.25, p > .05$. Specifically, in cases of failure, the outcome was not found to affect attributions to internal or external causes, $\chi^2(1, N = 21) = 2.43, p > .05$. More specifically, there was not a significant effect of the outcome on attributions to ability, $t(19) = 0.08, p > .05$, effort, $t(19) = -0.16, p > .05$, task difficulty, $t(19) = 0.72, p > .05$, or luck, $t(19) = -0.40, p > .05$, in cases of failure.

Future expectancy is the second facet of this hypothesis. In line with this study’s finding that stability did not have an effect on future expectancy, the locus of causality was not found to effect future expectancy either, $t(47) = -0.78, p > .05$.

**Investigational Analyses**

The first investigational analysis looked at BMI influences the attributions people make. A discriminant analysis was used to assess this relationship. In this analysis it was found that BMI did not affect the attributions people made, $\Lambda = 0.98, F(3, 44) = 0.37, p > .05$.

Table 4 (next page) presents the mean attribution and performance scores for success and failure groups that received either criterion- or norm-referenced feedback. Bear in mind that the sample sizes for the success and failure groups within the two methods of reporting are unequal due to the fact that participants were assigned to these two groups based on self-reported baselines for success.
### Table 4

**Success and Failure Groups under Criterion- and Norm-Referenced Conditions**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Criterion-Referenced</th>
<th>Norm-Referenced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success</td>
<td>Failure</td>
</tr>
<tr>
<td></td>
<td>(n = 13)</td>
<td>(n = 11)</td>
</tr>
<tr>
<td><strong>Raw score performance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>𝑀</td>
<td>27.54</td>
<td>19.00**</td>
</tr>
<tr>
<td>𝑆𝐷</td>
<td>1.21</td>
<td>1.32</td>
</tr>
<tr>
<td><strong>Ability attribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>𝑀</td>
<td>1.54</td>
<td>1.64</td>
</tr>
<tr>
<td>𝑆𝐷</td>
<td>0.28</td>
<td>0.31</td>
</tr>
<tr>
<td><strong>Effort attribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>𝑀</td>
<td>2.00</td>
<td>2.55</td>
</tr>
<tr>
<td>𝑆𝐷</td>
<td>0.27</td>
<td>0.30</td>
</tr>
<tr>
<td><strong>Luck attribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>𝑀</td>
<td>1.39</td>
<td>.55*</td>
</tr>
<tr>
<td>𝑆𝐷</td>
<td>0.29</td>
<td>0.31</td>
</tr>
<tr>
<td><strong>Task attribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>𝑀</td>
<td>1.08</td>
<td>1.27</td>
</tr>
<tr>
<td>𝑆𝐷</td>
<td>0.23</td>
<td>0.25</td>
</tr>
</tbody>
</table>

*Difference between success and failure groups significant at .05 level

**Difference between success and failure groups significant at .01 level
Table 4 reveals the average number of times each attribution-linked statement was checked on the questionnaire in each of the conditions. The scores could range from 0-3. For example, the score of 2.27 for effort under success with norm-referenced feedback means that the effort statement, “I did well on this test because I tried hard on each question,” was checked an average of 2.27 times.

The table shows two significant differences, the most notable of which is between the use of luck as an attribution under success and failure conditions when criterion-referenced feedback is used. In this case, luck was used significantly more as an explanation for success than it was for failure. Other than this, attributions to the four different causes were not influenced by a success or failure outcome, $\chi^2(3, N = 49) = 2.14, p > .05$. Attributions to the four causes were also not influenced by the method of reporting, $\chi^2(3, N = 49) = 2.68, p > .05$.

There was also a significant difference between the scores of success and failure groups under criterion-referenced feedback. When this difference was controlled for, the outcome was found to have a significant effect on attributions to ability, $F(1, 44) = 5.29, p < .05$. However, this effect was in the reverse direction as was hypothesized. There was more endorsement of the ability-linked statements following failure ($M = 1.90, SD = .24$) than success ($M = 1.09, SD = .21$).

Table 4 also illustrates that effort was the predominate attribution in all four categories, especially following failure under both feedback methods. When participants were categorized into one of the four attributions, 63.3% made attributions to effort with ability second at 20.4%. While attributions to ability and effort were anticipated to
predominate, effort was not expected to be as pervasive as it turned out to be. A possible reason for the prevalence of effort attributions is examined in the next section.

Discussion

None of the hypotheses were supported by the data. In fact, the reverse occurred in the case of luck attributions when using criterion-referenced feedback and ability attributions when performance was controlled for. The study also did not reproduce previous research findings. For instance, stability has been strongly linked to future expectancy (see Weiner, 1985), but this was not found in the current study. Also, the self-serving bias was not shown to play a role in attribution making to internal and external causes as it has in the literature (see, for example, Mezulis et al., 2004). The fact that the study failed to reproduce previous findings in the literature points to possible methodological flaws and limitations. These are addressed in the following sections.

*Expectancy-Valence Theory: Level of Motivation*

A primary limitation and concern of the current study is the level of motivation the participants had in completing the study. The standardized test questions were formatted and timed to take approximately thirty-five to forty minutes. However, some of the participants completed the items in less than twenty minutes, with some finishing in less than ten minutes. In order to provide an explanation for this, Expectancy-Valence Theory will be briefly discussed. The theory is generally associated with Vroom, Lawler, Hackman, and Porter (Altman, Valenzi, & Hodges, 1985). The theory states that a person’s motivation to achieve a goal depends on the value they place on achieving that goal (its valence) and the person’s belief that a certain action will lead to a certain outcome (its expectancy) (Feather & Newton, 1982). In other words, “people are motivated to do something if they see something
in it for themselves” (Altman, et al., 1985, p. 68). This theory portrays motivation as an association between three components (Lawler, 1969): expectancy, instrumentality, and valence. *Expectancy* represents “effort-reward probability” (Lawler, 1969, p. 161). For example, a student might believe that a lot of studying (i.e. working harder) will result in a higher grade on the test. This represents the belief that a particular behavior will result in a relatively immediate, measurable outcome or reward. *Instrumentality* represents the relationship between an immediate outcome, and an ultimate personal outcome. In this case, the student may believe that a higher grade will result in more praise and positive comments from their teachers, students, and/or peers. *Valence* represents the strength of preference that, in this case, the student has for the higher grade or higher praise outcome. It also refers to how highly the ultimate result is valued. The relationship between these three factors is multiplicative, meaning that if any factor is only slightly present or nonexistent, there will be little to no motivation present.

In the current study, participants may not have had the requisite motivation to adequately complete the standardized test questions. To illustrate this point, if a student wants to get one of the highest scores on the test, then getting a high score has high valence for that student. If the student believes that high performance will result in a high score, then the student has a high expectancy. However, if the student is not sure about how they will rank relative to the other students because they are unsure how much effort is required to get one of the highest scores, then the student has low instrumentality, and the student will not be motivated to perform better. This example assumes that the student wants to get a relatively high score, which may not necessarily be the case. If the student places little value on getting a high score, then the task as a whole has low valence. Nathawat, Singh, and Singh (1997)
differentiated participants on high versus low valence for achievement (which they termed “need for achievement”). The study found differences in the types of attributions the participants made with the low need for achievement group attributing negative outcomes to internal and stable (as well as global) factors. The example of the student who wants a high score also provides an illustration of how norm-referenced testing may inhibit motivation. Even if the student values getting a high score and understands that higher effort will result in a higher score, they still may not possess optimal motivation in a norm-referenced testing situation because they are unsure how much effort will be put forth by the other students. Therefore, they do not know how much effort they need to put into the task to reach their goal.

It was originally planned in the current study that the students’ actual test responses to a state-mandated test would be scored and these scores would be presented to the students. Unlike this intended method, the task involved in the present study was not a real, ongoing life event that held some importance to the students involved. However, the state-mandated tests have more importance to the younger students and it can be reasonably assumed that the younger students would have been more motivated in this context relative to the college students taking an arbitrary test to which they may not have assigned any importance. As was the original intent of this study, future studies should use a salient event that would have been experienced by the participants even if the study was not conducted. Such events could be a class exam, state-mandated standardized test, performance on a project, and so forth.

Causal Attributions

A second limitation is the narrow range of attributional choices participants were presented with. The current study primarily focused the causal dimensions of locus of
control and stability. Moreover, the study utilized Weiner’s (1974) four important factors affecting attributions for achievement: ability, effort, task difficulty, and luck. While these four general causes in combination cover all of the dimensions, they are still limiting. Other studies have utilized a wider array of possible attributions and have even sometimes allowed for open-ended responses that are later coded into causal dimensions. Weiner, Russell, and Lerman (1979), for instance, used six causes (ability, unstable effort, stable effort, personality, other people, and luck) as the factors they asked the participants to use in assigning cause to twelve achievement contexts. Another study that asked students to answer closed-ended items utilized even more attributional causes (Elig & Frieze, 1979). These included: high/low general intelligence, task ease/difficulty, good/bad luck, high/low unusual effort, high/low stable effort, high/low task interest, good/bad mood, and high/low motivation. Flammer and Schmid (2003) asked children to provide open-ended reasons for a success or failure scenario that was read to them. For each child a maximum of three responses was coded and differentiated among fifteen different broad categories. Such categories included, but were not limited to, ability, experience, effort, motivation, difficulty, social facilitation, copying from a neighbor, chance, and health. Several of the broader categories were further discriminated into smaller subcategories. For instance, effort was split into effort during the task, concentration, external distraction, and mental distraction. Using this approach, the researchers generated thirty-seven categories and subcategories for the causal attributions given by the children.

This approach encapsulates more possible explanations for a success or failure outcome than the standard four-attribute approach utilized in this study. The original intent of using this approach was to simplify the questionnaire for the 4th- and 5th-graders and to
make the questionnaire more efficient and straightforward. However, the attribution questionnaire format and content was maintained once the population was switched over to college students due to time constraints in formatting a new questionnaire as well as concerns surrounding survey fatigue. Future studies should at least allow participants to choose from much more differentiated causes if not allowed to generate their own causes that are later coded into the appropriate causal dimensions.

Another concern with regard to causal dimensions is the view of effort as an unstable cause. The purpose of the narrow view of effort imparted in this study was to simplify the range of attributional choices for the originally intended population. However, as evidenced in Table 1, there are two types of effort: immediate and typical. Typical effort is considered stable and immediate effort is considered unstable. If effort is viewed as a stable cause as well, then criterion-referenced results could be hypothesized to mediate higher expectancy in the future in addition to norm-referenced results. While this is not an immediate concern of this study as neither type of feedback was found to influence expectations, this is something that future studies should consider.

**Sample Size**

A final, and more evident, limitation of the current study is its low sample size. Most of the statistical tests were done with 56 participants, but the tests that broke down the participants into two subgroups (stable versus unstable) typically had sample sizes less than twenty in each group. The sample size was cut down even further when the success and failure groups were utilized whereby the largest sample size was for the success outcome under norm-referenced testing (\(n = 15\)).
In spite of the nonsignificant findings of the current study, there are other directions this vein of research could go. One possible route is addressing personal dispositions.

*Learned Helpless versus Mastery-Oriented Dispositions*

While the current study differentiated primarily on the basis of criterion- and norm-referenced results and how these affect attributions, Dweck (1986) makes a primary distinction between students that are *mastery-oriented* and those that are *learned-helpless oriented*, and how these general dispositions affect attributions. These two orientations are not seen as being on two ends of a single spectrum, rather as two distinct, complex categories of academic attitudes and behaviors that students have been shown to exhibit. Mastery-oriented students tend to adopt learning goals, which are set to increase competence (Dweck, 1986; Elliot & Dweck, 1988). These students are likely to seek challenges because they believe they will lead to great competence. Mastery-oriented students also increase their effort when they are faced with a failure outcome (Dweck, 1986; Elliot & Dweck, 1988). In terms of the attributions they make, mastery-oriented students tend to attribute their successes to high ability and their failures to events outside of themselves.

In contrast to learning goals, performance goals are set by students who attempt to gain favorable judgments or to avoid criticism. Students who utilize these goals are more likely to avoid challenges and tend to respond to failure with a sense of learned helplessness (Dweck & Leggett, 1988; Elliott & Dweck, 1988). Learned helplessness refers to the experience-based expectation that’s one’s actions cannot produce success (Dweck, 1975). Learned-helpless students are more concerned about appearing competent and view mistakes as sign of their inadequacies. In this sense, students who use performance goals are more likely to make attributions to ability in failure situations. Also, learned-helpless students tend
to view ability as a more stable entity than do mastery-oriented students. Because of the apparent advantages learning goals, Heyman and Dweck (1992) recommend that students are encouraged to focus primarily on these. They also argue that performance goals should play a smaller role by allowing children to enjoy recognition for their accomplishments, but this should not become a superseding concern. In order to reach this end, they suggest that teachers focus on learning rather than subjective comparisons when encouraging students. Future studies can look out how criterion- and norm-referenced results differentially affect students who have adopted a mastery-oriented or learned-helpless learning attitude. This form of research could provide useful information about how to effectively intervene with these students and maximally motivate them, especially ones with learned-helpless orientations. This is especially relevant in a time that is seeing standardized tests becoming more prevalent in schools. In other words, research can be done on how these tests are affecting students with each of the learning orientations. Also, as of now, the bulk of the research on this topic has been done with school-aged children as the population of interest. Therefore, another vein of research could examine these two orientations in college students.

The current study focused on situational cues and how they affect attributions and future expectancy. By introducing Dweck’s line of research, dispositional factors could be addressed as well.

Conclusions

The purpose of the current study was to examine a relatively overlooked aspect of standardized testing—the presentation of results. The aim of the study was to take a very practical approach, within theoretical frameworks, to examine this topic by examining how the two types of feedback mediate attributions and future expectancy. In spite of the current
student’s nonsignificant findings, this vein of research is very important and should be explored further. This is because of the widespread nature of standardized testing and the high levels of feedback students receive not only from standardized tests, but from their teachers, parents, and peers. Future studies need to address the limitations of this study and possibly take the general ideas presented and apply them to similar areas of research. The simple manipulation of categorizing feedback into the fairly comprehensive categories of norm- and criterion-referenced can be easily utilized to examine how the style of feedback differentially affects other constructs relevant to the educational domain such as achievement motivation, academic self-efficacy, classroom atmosphere (i.e. cooperative versus competitive) and even emotional affect. This manipulation can also be extended to examine feedback in other settings such as the workplace or sporting environments. This type of research has the potential of offering very practical implications for how feedback affects different constructs in many contexts and environments.
References


Butler, R. (1987). Task-involving and ego-involving properties of evaluation: effects of
different feedback conditions on motivational perceptions, interest, and performance.
*Journal of Educational Psychology, 79*(4), 474-482.

Chan, L. K., & Moore, P. J. (2006). Development of attributional beliefs and strategic
knowledge in years 5-9: a longitudinal analysis. *Educational Psychology, 26*(2), 161-
185.

Approaches to learning, attributions, goal orientations, and achievement: a
developmental perspective. Paper presented at the Australian Association for
Research in Education conference, Geelong, Australia.

performance: national study of kindergarteners and first-graders. *Obesity Research,
12*(1), 58-68.


Dweck, C. S. (1975). The role of expectations and attributions in the alleviation of learned

41*(10), 1040-1048.


### Appendix A

**Criterion-Referenced Reporting Sample**

<table>
<thead>
<tr>
<th>Participant</th>
<th>Score</th>
<th>Percentage</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant A</td>
<td>23/41</td>
<td>56%</td>
<td>Not Successful</td>
</tr>
<tr>
<td>Participant C</td>
<td>19/41</td>
<td>46%</td>
<td>Not Successful</td>
</tr>
<tr>
<td>Participant E</td>
<td>37/41</td>
<td>90%</td>
<td>Successful</td>
</tr>
</tbody>
</table>
## Appendix B

Norm-Referenced Reporting Sample

<table>
<thead>
<tr>
<th>Participant</th>
<th>Rank</th>
<th>Percentile</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>19&lt;sup&gt;th&lt;/sup&gt; out of 44*</td>
<td>57&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>Successful</td>
</tr>
<tr>
<td>D</td>
<td>6&lt;sup&gt;th&lt;/sup&gt; out of 54*</td>
<td>89&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>Successful</td>
</tr>
<tr>
<td>F</td>
<td>41&lt;sup&gt;st&lt;/sup&gt; out of 57*</td>
<td>28&lt;sup&gt;th&lt;/sup&gt; percentile</td>
<td>Not Successful</td>
</tr>
</tbody>
</table>

*Number represents how many participants had completed the first time-point when the participant got their score.
Appendix C

Sample SAT Questions

Sentence Completions

Instructions: Choose the answer that contains the words that make the most sense in the sentence.

There is no doubt that Larry is a genuine ------: he excels at telling stories that fascinate his listeners.

A) Braggart B) Dilettante C) Pilferer D) Prevaricator E) Raconteur

Improving Sentences

Instructions: The following sentences test correctness and effectiveness of expression. Part of each sentence or the entire sentence is in parentheses; beneath each sentence are five ways of phrasing the parenthesized material. Choice A repeats the original phrasing; the other four choices are different. If you think the original phrasing produces a better sentence than any of the alternatives, select choice A; if not, select one of the other choices.

(Looking up) from the base of the mountain, the trail seemed more treacherous than it really was.

A) Looking up B) While looking up C) By looking up D) Viewing E) Viewed
Multiple-Choice Mathematics

Instructions: Answer the following multiple-choice mathematics questions by clicking on the most accurate response. Scratch paper and calculators are permitted.

A jar contains four blue marbles and two green marbles. Without looking, two marbles are drawn from the jar. What is the probability that two marbles of the same color will be selected?

A) 1/9  B) 2/9  C) 1/3  D) 7/15  E) 7/9

Open-Ended Mathematics

Instructions: Please attempt to provide the correct response in the space provided. Fraction and decimal formats are acceptable.

Three parallel lines in a plane are intersected by a fourth line, forming twelve angles. If one of the angles has measure 28°, how many of the other eleven angles have measure 28°?

Answer: 5
Appendix D

Follow-Up Questionnaire

The following statements ask you to choose the option that best describes why you think you were *successful* (italics added) on the test. Please check only one of the two options that are available for each statement, with the one you check being the most accurate response. Answer each question independently of the other questions; as if the other questions do not exist.

1. I was successful on the test mainly because:
   _____ I tried hard on each question   OR   _____ I was lucky

2. I was successful on the test mainly because:
   _____ I was lucky   OR   _____ The test was easy

3. I was successful on the test mainly because:
   _____ I tried hard on each question   OR   _____ The test was easy

4. I was successful on the test mainly because:
   _____ I am good at these tests   OR   _____ The test was easy

5. I was successful on the test mainly because:
   _____ I was lucky   OR   _____ I am good at these tests

6. I was successful on the test mainly because:
   _____ I tried hard on each question   OR   _____ I am good at these tests

7. How well do you expect to do on a similar test in the future?
   o Much better than this time
   o Slightly better than this time
   o The same as this time
   o Slightly worse than this time
   o Much worse than this time
## Appendix E

Table 1

*Causes of Success and Failure and their Corresponding Dimensions*

<table>
<thead>
<tr>
<th></th>
<th>Internal</th>
<th></th>
<th>Internal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Controllability</td>
<td>Stable</td>
<td>Unstable</td>
<td>Stable</td>
<td>Unstable</td>
</tr>
<tr>
<td>Uncontrollable</td>
<td>Ability</td>
<td>Mood</td>
<td>Task difficulty</td>
<td>Luck</td>
</tr>
<tr>
<td>Controllable</td>
<td>Typical effort</td>
<td>Immediate effort</td>
<td>Teacher bias</td>
<td>Unusual help from others</td>
</tr>
</tbody>
</table>
Table 2

Criterion- and Norm-Referenced Test Differences

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Criterion-Referenced Tests</th>
<th>Norm-Referenced Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>To determine whether each student has achieved specific skills or concepts.</td>
<td>To rank each student with respect to the achievement of others in broad areas of knowledge.</td>
</tr>
<tr>
<td></td>
<td>To find out how much students know before instruction begins and after it has finished.</td>
<td>To discriminate between high and low achievers.</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td>Measures specific skills which make up a designated curriculum. These skills are identified by teachers and curriculum experts.</td>
<td>Measures broad skill areas sampled from a variety of textbooks, syllabi, and the judgments of curriculum experts.</td>
</tr>
<tr>
<td></td>
<td>Each skill is tested by at least four items in order to obtain an adequate sample of student performance and to minimize the effect of guessing.</td>
<td>Each skill is usually tested by less than four items.</td>
</tr>
<tr>
<td><strong>Item Characteristics</strong></td>
<td>The items which test any given skill are similar in difficulty.</td>
<td>Items vary in difficulty.</td>
</tr>
<tr>
<td></td>
<td>Each individual is compared with a preset standard for acceptable achievement. The performance of other examinees is irrelevant.</td>
<td>Each individual is compared with other students and assigned a score--usually expressed as a percentile, a grade equivalent score, or a stanine.</td>
</tr>
<tr>
<td><strong>Score Interpretation</strong></td>
<td>A student’s score is usually expressed as a percentage.</td>
<td>Student achievement is reported for broad skill areas, although some norm-referenced tests do report student achievement for individual skills.</td>
</tr>
<tr>
<td></td>
<td>Student achievement is reported for individual skills.</td>
<td></td>
</tr>
</tbody>
</table>

70
Table 3

*Bivariate Correlations among SAT Scores*

<table>
<thead>
<tr>
<th></th>
<th>2.</th>
<th>3.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Math Scores</td>
<td>.739**</td>
<td>.937**</td>
</tr>
<tr>
<td>2. Verbal Scores</td>
<td></td>
<td>.928**</td>
</tr>
<tr>
<td>3. Total Scores</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Significant at the .01 level
Table 4

*Success and Failure Groups under Criterion- and Norm-Referenced Conditions*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Criterion-Referenced</th>
<th>Norm-Referenced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success (n = 13)</td>
<td>Failure (n = 11)</td>
</tr>
<tr>
<td>Raw score performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>27.54</td>
<td>19.00**</td>
</tr>
<tr>
<td>SD</td>
<td>1.21</td>
<td>1.32</td>
</tr>
<tr>
<td>Ability attribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1.54</td>
<td>1.64</td>
</tr>
<tr>
<td>SD</td>
<td>.28</td>
<td>.31</td>
</tr>
<tr>
<td>Effort attribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>2.00</td>
<td>2.55</td>
</tr>
<tr>
<td>SD</td>
<td>.27</td>
<td>.30</td>
</tr>
<tr>
<td>Luck attribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1.39</td>
<td>.55*</td>
</tr>
<tr>
<td>SD</td>
<td>.29</td>
<td>.31</td>
</tr>
<tr>
<td>Task attribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1.08</td>
<td>1.27</td>
</tr>
<tr>
<td>SD</td>
<td>.23</td>
<td>.25</td>
</tr>
</tbody>
</table>

*Difference between success and failure groups significant at .05 level*

**Difference between success and failure groups significant at .01 level*