AUGMENTING THE ACHIEVEMENT GOAL QUESTIONNAIRE-REVISED

A Thesis Presented to the
Faculty of the College of Education
University of Houston

In Partial Fulfillment
of the Requirements for the Degree

Master of Education

by

Justin Neil deLeon Young

May, 2012

AUGMENTING THE ACHIEVEMENT GOAL QUESTIONNAIRE-REVISED

A Thesis for the Degree

Master of Education

by

Justin Neil deLeon Young

Approved by Thesis Committee:

____

Dr. Richard Olenchak, Chairperson

____

Dr. Christopher A. Wolters, Committee Member

____

Dr. Shirley Yu, Committee Member

____

Dr. Robert H. McPherson, Dean
College of Education

May, 2012

AUGMENTING THE ACHIEVEMENT GOAL QUESTIONNAIRE-REVISED

An Abstract
of a
Thesis Presented to the
Faculty of the College of Education
University of Houston

In Partial Fulfillment
of the Requirements for the Degree

Master of Education

by

Justin Neil deLeon Young

May, 2012

Young, Justin Neil L. “Augmenting the Achievement Goal Questionnaire-Revised.” Unpublished Masters of Education Thesis, University of Houston, May 2012.

Abstract

Items drafted to measure *mastery-approach*, *mastery-avoidance*, *performance-approach*, and *performance-avoidance* goals were administered with Elliot and Murayama’s (2008) AGQ-Revised (AGQ-R) to 300 students to see how responses fit the 2 x 2 factor model of achievement goal theory, and how the reliability as well as the variance of the AGQ-R scores were affected by the drafted items.

Three Multiple-Indicator Correlated Trait-Correlated Method models (MI CT-CM) of the 2 x 2 achievement goal theory were examined using scores from the AGQ-R, drafted items, and the AGQ-R plus drafted items. No model reached non-significance, but the disturbed MI CT-CM model with the AGQ-R exhibited the best model fit. This model may not be viable because of Heywood cases. Drafted items were also examined to see how the AGQ-R might be improved.

Reliability of the *mastery-approach* and *performance-approach* responses significantly increased. Reliability of the *performance-avoidance* responses significantly decreased. Variance in every subscale, except *performance-approach*, significantly decreased.

TABLE OF CONTENTS

LIST OF TABLES

The No Child Left Behind Act (2001) was passed in order to “ensure that all children have a fair, equal, and significant opportunity to obtain a high quality education and reach, at a minimum proficiency on challenging state academic achievement standards and state assessments” (Statement of Purpose, para. 1). It recognizes that some children do have the opportunity to obtain a high quality education and endorses “closing the achievement gap between high- and low-performing children”(Statement of Purpose, para. 1). Different reasons can be attributed to this achievement gap (cf. Levin, 2007, McCaslin, 2006, Mulvenon, Wang, McKenzie, & Airola, 2006); however, one seems particularly relevant to motivational science. McCaslin (2006) posited one reason for the achievement gap is that some students possess knowledge of motivation and self-regulation while other students do not. The implication is that the study and investigation of motivation is important and research on academic motivation should continue because of its pragmatic implications.

Motivation has been researched for quite some time and continues to change in regards to academic achievement. Atkinson (1983) developed an expectancy-value framework for achievement motivation using Murray’s definition of need for achievement. Under Atkinson’s original framework, motivation is posited to be a product of motive and expectancy. Raynor (1969) incorporated incentive into the theory to explain why people engage in activities that have little to no value to them. This seemed to make expectancy-value theory more pragmatic, but both versions of the expectancy-value framework view motivation as unidimensional. However, this is problematic because Jackson, Ahmed, and Heapy (1976) have demonstrated that motivation is multidimensional. There are variations of Atkinson’s original expectancy-value theory which are multi-dimensional (cf. Eccles, 1983;

Pintrich & DeGroot, 1990), but this research used achievement goal theory, which is multi-dimensional, to explain motivation.

Achievement goal theory looks at the quality of a student's motivation, and an achievement goal can be viewed as essentially the reason why a student chooses to engage or not engage in a task. Initially only two types of goals were researched: *mastery* goals and *performance* goals (Ames, 1992). Students chose to engage in a task because they either wanted to master the task or perform the task better than their peers. However, researchers have come to understand that these goals can be partitioned and viewed in a different framework. Elliot and Harackiewicz (1996), for example, used valence to divide *performance* goals into *performance-approach* and *performance-avoidance* goals. Students who want to perform a task better than their peers are said to have adopted a *performance-approach* goal, whereas students who do not want to perform a task worse than their peers are said to have adopted a *performance-avoidance* goal.

Since then, Elliot (1997) proposed the 2 x 2 factor model of achievement goal theory that incorporates the valence of both *mastery* and *performance* goals. Valence divided *mastery* goals into *mastery-approach* and *mastery-avoidance* goals in the same manner it divided *performance* goals. Students who want to master, or understand, a task are said to have adopted a *mastery-approach* goal, whereas students who want avoid not mastering a task are said to have adopted a *mastery-avoidance* goal. This framework, which incorporates *mastery-approach, mastery-avoidance, performance-approach*, and *performance-avoidance* goals, is referred to as the 2 x 2 factor model of achievement goal theory and has been used in empirical research on motivation. However, some of the findings from the empirical research using achievement goal theory contradict each other.

Purpose

For further understanding of achievement goal theory, this research used structural equation modeling to investigate the extent to which a self-report questionnaire, the Achievement Goal Questionnaire-Revised (AGQ-R; Elliot & Murayama, 2008), fit the 2 x 2 factor model of achievement goal theory. Additionally, items were also drafted to in an attempt to increase the variance of AGQ-R scores. Muis, Winne, and Edwards (2009) examined the psychometric properties of instruments that measured achievement goals and noted the instruments had difficulty distinguishing between students. They surmised this could have been the case because of the limited variance of the scores, which is why items were drafted for this research.

REVIEW OF RELATED LITERATURE

Pintrich (2003) contended that research on motivation “can be and should be approached from a scientific perspective” (p. 667). As such, empirical evidence and assessment is required to test principles of achievement goal theory. There are different methods for measuring motivation. Phenomenological, neuropsychological, and behavioral methods can be used to collect evidence, but the self-report method is the most widely used (Fulmer & Frijters, 2009).

The Achievement Goal Questionnaire (AGQ; Elliot, 1997) and Patterns of Adaptive Learning Scales (PALS; Midgley et al., 2000) traditionally have been used to correlate dimensions of motivation with behavior and cognitive abilities. For example, Elliot and McGregor (1999) documented that a mastery orientation is a positive predictor of long-term retention using the AGQ. Anderman, Maehr, and Midgley (1999) documented a decline in motivation following school transitions using the PALS.

The AGQ provides scores with good reliability. Elliot and McGregor (1999) used the trichotomous model of achievement goal theory and reported coefficient alphas between .84 and .92 for *mastery*, *performance-approach*, and *performance-avoidance* goals. Finny, Pieper, and Barron (2004) examined the psychometric properties of the AGQ in a general academic context and reported coefficient alphas between .675 and .876 for the 2 x 2 factor model.

However, there are problems with the AGQ. Elliot and Murayama (2008) delineated seven concerns with the AGQ and created the AGQ-Revised (AGQ-R) to address these problems. Their concerns are described in the following three paragraphs, and the table in Appendix A summarizes the changes Elliot and Murayama made to the AGQ as well as includes examples of the original item and modified item.

First, goals are supposed to influence behavior, but the stems in the AGQ reflected values as opposed to actual behavior. Elliot and Murayama (2008) changed the stems in the items to include the word “goal”, “aim”, or “striving.” Second, the AGQ items combined the goal with the reason for the goal, and this confounded the measurement of the goal. This is why Elliot and Murayama removed the reason for the goal from the items to address that issue. Third, references to grades actually applied to both mastery and performance contexts. That is misleading and corrected in the AGQ-R by removing references to grades. Fourth, normative measures were stressed in items measuring *performance-approach* goals, but they were not even mentioned in items measuring *performance-avoidance* goals. The AGQ-R explicitly included normative references in the items measuring *performance-avoidance* goals to exclude a bias between *performance-approach* items and *performance-avoidance* items. The fifth concern Elliot and Murayama had, was that several items in the AGQ included an affective component. Since the focus of achievement goals is not the affective domain, the affective components were removed from items for the AGQ-R.

Sixth, an item on the AGQ did not allow the adoption of multiple goals because it used the word "just" to exclude the adoption of other goals. A *performance-avoidance* item read, "I just want to avoid doing poorly in this class", and an endorsement of this item theoretically precluded a participant from adopting goals other than *performance-avoidance*. Elliot and Murayama (2008) removed the word "just" from the item to allow for the adoption of different or multiple goals.

The final concern that Elliot and Murayama (2008) expressed with the AGQ is that it possessed items which highlighted extreme groups by including the word "most." In other words, Elliot and Murayama believed an extreme perceived competence tends to promote either *performance-approach* or *performance-avoidance* goals. If a student performed relatively worse than *most* of his/her peers, then that student is believed to be likely to adopt *performance-avoidance* goals because the student performed worse than *most* of the other students. However, Elliot and Murayama asserted that a student could still adopt a *performance-avoidance* goal even if he/she did not perform worse than *most* of the other students. As such, the word "most" was removed from the original items.

The improvements make sense from a nomological standpoint and the coefficient alpha for scores is even higher on the revised *performance-avoidance* scale. However, Elliot and Murayama (2008) only rewrote the items in their revision of the AGQ, which means the AGQ-R still has the same disadvantage the AGQ, and even PALS, does. Specifically, both the AGQ and PALS have short scales (fewer than seven items). Brief assessments may not always be accurate (Maloney, Grawitch, & Barber, 2011). There is a *bandwidth-fidelity* *tradeoff* which suggests the accuracy of an assessment is affected by the assessment's length (Ones & Viswesvaran, 1996). The AGQ-R has only 12 items that cover four factors, which means each factor has only three items. Accurate measurement is exacerbated by brief assessments when the construct is complex.

Inconsistencies in Research

Mixed and conflicting results in achievement goal theory demonstrate its complexity, and additional improvements can be made from an assessment standpoint by adding items to obtain a more precise conception of achievement goal motivation. As a whole, the following research in this section highlights some of the inconsistencies in achievement goal theory as well as rationalizes studies like this one which investigate the complexities of achievement goal theory.

Middleton and Midgley (1997) conducted research that correlated three factors of goal theory with adaptive behaviors (e.g., academic efficacy and self-regulated learning strategies) and maladaptive behaviors (e.g., test anxiety and avoiding help-seeking). They investigated *performance-approach*, *performance-avoidance*, and *task* goals. *Task* was operationalized in this study as developing ability. *Mastery* goals and *task* goals are equivalent (cf. Ames, 1992). Their cross-sectional correlational study of elementary school students showed that p*erformance-approach* and *performance-avoidance* goals were correlated with maladaptive behaviors including test anxiety. *Performance-approach* goals accounted for less than 10% of the variance in test anxiety (r = .32).

Elliot and McGregor (1999) studied the relation between *performance* goals and test anxiety as well. They viewed test anxiety as manifesting in traits (i.e., overall/general anxiety) as well as states (i.e., transitory anxiety), and they measured state anxiety in undergraduate students immediately after they completed an exam. These students' achievement goals were also assessed three weeks prior to that. Simultaneous multiple regression analyses of students' achievement goals and state anxiety after completing their exam provided evidence that the relation between *performance-approach* goals and exam performance is independent of test anxiety. In other words, Elliot and McGregor's findings contradicted Middleton and Midgley's (1997) findings. However, *performance-approach* goals are still associated with test anxiety and other maladaptive behaviors (Linnenbrink, 2005).

In a review of the literature, Midgley, Kaplan, and Middleton (2001) cited studies which provide evidence that *performance* goals are unrelated to deep processing. Midgley et al. also cite studies which provide evidence that *performance* orientations are positively related to deep processing. There are studies that provide evidence that *performance* goals are related to achievement outcomes (Midgley et al., 2001). There are also studies that provide evidence that *performance* goals are not related to achievement outcomes (Midgley et al., 2001). After assessing achievement goals in elementary school students, Linnenbrook (2005) demonstrated a negative relation between *performance-approach* goals and achievement. This is in contrast to the positive relation between *performance-approach* goals and achievement Wolters, Yu, and Pintrich (1996) demonstrated after assessing middle school students' achievement goals and performance outcomes.

Church, Elliot, and Gable (2001) conducted research that examined how undergraduates' perceptions of the classroom environment affected the adoption of goal orientations and ultimately graded performance and other achievement outcomes. Correlations between *performance-approach* goals and achievement outcomes such as SAT scores and grades were weak and positive, but they were comparable to the correlations between *mastery* goals and achievement outcomes, which were also weak and positive. Conversely, the same study, in which Middleton and Midgley (1997) associated *performance-approach* goals with test anxiety, also provided evidence that *performance-approach* goals are negatively correlated with prior achievement.

It should be noted that Church et al. (2001) used the AGQ and that Middleton and Midgley (2001) used the PALS because a meta-analysis by Hulleman, Schrager, Bodmann, and Harackiewicz (2010) demonstrated that the instrument used (either the AGQ or the PALS) will affect correlations between achievement goals and educational outcomes. These differences can be accounted for by construct specification or operationalization.

The Importance of Variance

Muis et al. (2009) conducted Rasch-model analyses of the AGQ and PALS which indicated that the instruments cannot effectively measure actual individual differences. They observed that respondents are more likely to agree, or respond positively, with items that measure *approach* goals. This is what makes measuring individual differences difficult using the AGQ and PALS. Most, if not all, students agreed with the *approach* statements on the AGQ and PALS; therefore differentiating between individuals using these instruments is not practical. According to Muis et al., the AGQ and PALS actually measure achievement goals of the overall group and not the individuals. This can be corrected by adding "more cleverly worded items" to the instruments (Muis et al., 2009, p. 571), though any procedure that increases the variance of the scores is likely to enhance measurement of individual achievement goals. This project proposes to increase variance by adding more items to each subscale of the AGQ-R.

Many experiments investigate using only the three factor model of *mastery*, *performance-approach,* and *performance-avoidance* goals. Given that this was exploratory construct-oriented research, all four factors of achievement goal theory were examined. Part of the exploratory process is investigating how additional items function because examining the additional items allow for the understanding of *mastery-approach, mastery-avoidance, performance-approach*, and *performance-avoidance* goals to be assessed.

Hulleman, Schrager, Bodmann, and Harackiewicz (2010) performed a meta-analytic review of achievement goal measures and concluded that "achievement goal researchers are using the same label for different constructs" (p. 441). This may be evidence that *mastery-approach, mastery-avoidance, performance-approach*, and *performance-avoidance* goals are not fully understood. This project addresses that concern by investigating additional items to examine an understanding of how achievement goals are viewed.

Paradigm for Achievement Goals

Most achievement goal theorists view goals in terms of competency and valence. This is also the perspective of this research. A competency is a standard by which to assess goal attainment or success. The two competencies in achievement goal theory are *mastery* and *performance*. The *mastery* competence indicates that performance will be compared to a domain or individual standard. A domain-referenced standard is akin to "I want to be able to solve a system of equations." An individual-referenced, or ipsative, standard is akin to "I want to continue to be able to solve systems of equations."

However, achievement goal theorists make their own distinctions concerning how a *mastery* goal is defined and used. Midgley, Kaplan, and Middleton (2001) operationalize *mastery* as "developing competency and gaining understanding and insight" (p.77). They do not seem concerned with a sustainable aspect and seem to relegate that to a *performance* competency. However, Elliot and Murayama (2008) operationalize *mastery* to include an "intrapersonal" standard. This study included sustainability with the *mastery* competence, and goals may be domain-referenced or ipsative.

The *performance* competency indicates that performance will be compared to other individuals' performances. This means any goals that are norm-referenced are considered to have a *performance* competency. An example of a *performance* competency is "I want to do better than other students." As such, the distinction between *mastery* and *performance* competencies is that *mastery* goals reference prior individual performance or material from a domain and *performance* goals reference other individuals.

A valence is the direction in which one moves relative to competence or incompetence (Elliot, 1999). Competence is still the standard for assessing goal attainment, and incompetence is the standard for assessing lack of goal attainment. Valence is most easily seen in a *performance* competency like "I want to do better than other students." If one does better than other students, one has attained the goal. If one has not done better than other students, the goal was not attained. This results in the incompetency that states "I have done just as well or worse than other students."

Competencies are positive outcomes, which imply a positive valence indicates that one is moving toward a competency, or positive outcome. Thus the *performance-approach* goal is "I want to do better than other students." Incompetencies are negative outcomes, which imply a negative valence indicates that one is moving away from an incompetency, or negative outcome. Thus the *performance-avoidance* goal is "I do not want to do worse than other students."

To valence a *mastery* goal, what is opposite of learning needs to be determined. This study understands forgetting to be the opposite of learning, and the standard of incompetence for *mastery* will be forgetting. An individual who has adopted a *mastery-avoidance* goal will be concerned with not forgetting, or remembering, material. As such, a *mastery-avoidance* goal is akin to "I do not want to forget how to solve systems of equations." The inclusion of "not" is what distinguishes an *avoidance* goal from an *approach* goal.

The preceding framework essentially established four goals for exhibiting any behavior, according to achievement goal theory. Even though the AGQ was revised to eliminate values, goals may be seen as values. Elliot and Murayama (2008) operationalized a goal as "an aim that one is committed to that serves as a guide for future behavior" (p. 614) which is very clear. However, only four out of the 12 items on the AGQ-R indicate any kind of behavior. Items like "[m]y goal is to learn as much as possible" assess values because there is no indication of a behavior. Even though Fulmer and Frijters (2009) criticize instruments for assessing values as opposed to goals, motivational instruments should do both. An item like "[m]y goal is to learn as much as possible" indicates what type of goal an individual has. An items like "I am striving to understand the content of this course as thoroughly as possible" indicates the strength of a particular goal because the item looks at behavior. A stronger goal should be indicated by salient behavior since a goal is meant to guide behavior. Items were drafted with particular behaviors in mind because of this.

Each scale in the AGQ-R has only one item which looks at behavior; these items have the "I am striving" stem. For instance, two students may mark the same number on one of those items, but the responses may not mean that the strength of the goals is of the same magnitude. One student may study, and the other one may not study. It is reasonable to assume that the one that studies has a stronger goal, even though both marked the same number, because the goal guided future behavior.

Research Questions:

1. How does the AGQ-R with drafted items load onto the 2 x 2 factor model?
2. How do the drafted items affect the reliability of AGQ-R scores?
3. How do the drafted items affect the variances of AGQ-R scores?

The first research question was answered using structural equation modeling. It was hypothesized that the responses to the drafted items would fit the 2 x 2 factor model better than the combined responses of the AGQ-R and drafted items combined. However, it was also hypothesized that the responses to the AGQ-R would fit the 2 x 2 factor model better than the responses to the drafted items. The second and third research questions were answered by calculating the statistic (i.e., coefficient alpha and variance) for the AGQ-R subscales without the drafted items and with the drafted items. Additionally, bootstrapping was used to determine significant differences. It was hypothesized that both the reliability and variance of the AGQ-R subscale responses would increase by adding the drafted items.

METHOD

Participants

A total of 300 undergraduate students (263 female, 34 male, and 3 unidentified) from the University of Houston main campus (UH) and downtown campus (UHD) completed the AGQ-R and the drafted items. Participants either received academic credit for a course in which they were enrolled or candy of their choice. A total of 297 students reported their ethnicity; responses included African-American (n = 43, 15%), Asian (n = 83, 28%), Caucasian (n = 74, 25%), Hispanic (n = 71, 24%), Native American (n = 1, < 1%), and other (n = 25, 8%).

Measures

The instrument used in this research was a self-report survey, which consisted of two parts. The first part of the survey AGQ-R was modified to better suit this study (i.e., courses and academics were referenced in general as opposed to specific classes) and increase the variance. The second part of the survey consisted of items drafted within the 2 x 2 achievement goal theory framework. The drafted items followed a pattern which is similar to the guidelines Elliot and Murayama (2008) used to make changes to the AGQ. This pattern was established in order to map the items onto the different factors in the 2 x 2 achievement goal theory framework. The final instrument consisted of 64 items and is described in more detail below.

The first 12 items were taken from Elliot and Murayama's (2008) AGQ-R. They used a 7 point Likert scale, but Muis et al. (2009) suggested using a 5 point Likert scale because the categorical ratings in the Rasch analysis were unordered. Ratings should have been similar to a normal distribution, but they were not. Muis et al. posited ratings may have been unordered because a 7 point Likert scale was used. Not every numerical alternative on the AGQ-R had a label, and this allowed for the possibility of students to misinterpret them, or interpret them differently than one another. In an attempt to mitigate this possibility, each alternative for the AGQ-R and drafted items had an anchor statement. The anchor statements are described in the next paragraph, and the drafted items are described after that.

The instructions on the questionnaire asked students to respond to each statement by indicating how true each statement was on a typical day, and they were provided with the following anchors. A response of 1 would indicate that the item is always false. A response of 2 would indicate that the item is mostly false. A response of 3 would indicate that the response is sometimes true and sometime false. A response of 4 would indicate that the item is mostly true. A response of 5 would indicate that the item is always true.

Items were drafted by looking at staple tasks in an academic setting (i.e., studying and paying attention in class) and imagining how an individual with each achievement goal would behave. It is reasonable for an individual with a *mastery-approach* goal to "study to learn more about the material being taught." The reference to "material" is an indication that the item attempted to measure a *mastery* competence, and the word "learn" indicates that the item also attempted to measure an *approach* valence. To be consistent and clear on this point, all the drafted items intended to measure *mastery* goals reference "material." All the drafted items intended to assess *mastery-approach* goals use the word "learn" or "understand." The items hypothesized to measure *mastery-approach* goals can be found in Appendix B.

Similar to how a student who has adopted a *mastery-approach* goal would study in order to learn the material, it is reasonable for a student that has adopted a *mastery-avoidance* goal to "study in order to not forget the material being taught." Once again, the reference to "material" indicates the item attempted to measure a *mastery* competence, but the use of the word "not" indicates an *avoidance* valence. All the drafted items intended to assess *avoidance* goals use the word "not." The items hypothesized to measure *mastery-avoidance* goals can be found in Appendix C.

Using the same logic that established the pattern for hypothesized *mastery-approach* and *mastery-avoidance* goals, it is reasonable for that student who has adopted a *performance-approach* goal to "study in order to perform better than other students." The reference to "other students" is an indication that the item attempted to measure a *performance* competency, and the word "better" indicates that the item also intended to assess an *approach* valence. All the drafted items intended to measure the *performance* competence refer to "other students." All the drafted items intended to assess *performance-approach* goals use the word "better." The items hypothesized to measure *performance-approach* goals can be found in Appendix D.

For a student with a *performance-avoidance* goal, it is reasonable for that student to "study in order to not perform worse than other students." Once again, the reference to "other students" indicates the item attempted to measure a *performance* competence. The use of the word "not" indicates an attempt to measure the *avoidance* valence. The items hypothesized to measure *performance-avoidance* goals can be found in Appendix E.

Interestingly enough, all four goals may provide a student with a reason to study. However, not all students study. As such, to assess students who do not study, items were drafted within the *avoidance* valence because it seems reasonable that students with *mastery-avoidance* goals and *performance-avoidance* goals would be the ones to not study. A student with a *mastery-avoidance* goal is concerned with not forgetting and remembering, which means this student may not study because he/she has not forgotten, or remembers, the material (or at least the student may believe as much). A student with a *performance-avoidance* goal is concerned about doing as well as others, which means this student may not study because he/she knows (or believes) other students are not studying. In this sense, the student is not doing any worse than other students; he/she is doing exactly the same as other students. Individuals with *avoidance* goals may avoid certain academic tasks because they may believe that incompetency has already been avoided. The *mastery-avoidance* student believes he/she remembers the material; therefore incompetency has avoided and the task is irrelevant. The *performance-avoidance* student believes he/she can perform no worse than other students by acting as other students do.

Hulleman et al. (2010) indicated that a *performance-approach* goal may have an appearance component and normative component. It is not known whether the additional items which indicate an avoidance of behavior are actually the appearance component of *performance-approach* goals. Even if they are, there are two reasons for grouping them with *mastery-avoidance* and *performance-avoidance*. First, Hulleman et al. cited Urdan and Mestas (2006) when writing about the two components of *performance-approach*. Urdan and Mestas established the appearance and competition goals after interviewing individuals with “relatively high scores on a survey of measure of performance-avoidance goals” (p.354). As such, appearance and normative goals would be components of *performance-avoidance* goals. Second, grouping the additional items under *mastery-avoidance* and *performance-avoidance* follows the established paradigm of the word “not” intending to assess an *avoidance* goal.

The finished instrument consisted of Elliot and Murayama’s (2008) AGQ-R followed by the 52 drafted items in random order. Three data sets were produced from the survey. The first data set consisted of the responses from the AGQ-R. The second data set consisted of the responses to the drafted items. The third data set consisted of responses from the AGQ-R plus the drafted items. They will be referred to as the AGQ-R data, drafted data, and AGQR+ data respectively.

Procedures

Students were recruited from the UH Honors College during the spring semester, 2011 and UHD during the fall semester, 2011. Participants from the UH Honors College and UHD completed a paper version of the survey and received candy for their participation. To recruit students from the Honors College, announcements were made in courses offered by the UH Honors College and were sent out to a listserv of students enrolled in the Honors College. To recruit students from UHD, paper announcements were distributed in a few English courses and across their campus.

During the fall 2011 semester, an electronic version of the survey was made available to UH students who were required to participate in research. Students who chose to complete the survey received credit in the course that required them to participate in research. These students completed the survey online at a time and place of their choice.

Data Analysis

The AGQ-R data are similar to multitrait-multimethod (MTMM) data and require the use of structural equation modeling (SEM) as opposed to exploratory factor analysis to test the model. The 2 x 2 achievement goals can be decomposed into two separate trait factors (competency and valence) in the same manner that MTMM data can be decomposed into a trait factor and method factor (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). Furthermore, exploratory factor analysis cannot be performed using AGQ-R data because the correlations of the first order factors (achievement goals) can only be freely estimated with factors which they share a decomposition; correlations between first order factors that do not share a decomposition are restricted to zero (Marsh & Hocevar, 1988). For example, *performance-approach* goals can be decomposed into a *performance* competency and an *approach* valence. As such, the correlation of *performance-approach* goals and *mastery-approach* goals should be freely estimated because they have the *approach* valence in common. The correlation of *performance-approach* goals and *mastery-avoidance* goals should be restricted to 0 since they have different competencies *and* different valences. Exploratory factor analysis would not allow any correlations to be restricted to 0.

The first, and most important, step in SEM is to specify the model (Kline, 2011). Three models, based on Elliot and Murayama's (2008) study, were used. They used a multiple-indicator correlated trait-correlated method (MI CT-CM) model to test the 2 x 2 factor model. They "constrained paths from the same second factors to be equal and fixed the covariance between the mastery and performance factors to be 0" (Elliot & Murayama, 2008, p. 621). The rationale for the constraint was to stabilize the estimates. An unconstrained model was also tested in this study because there was not a nomological reason to institute Elliot and Murayama's constraints. In other words, no theory had been presented to justify constraining the second order factors to be equal. In this regard, allowing the second order factors to vary is equally justified and an alternative this study chose to explore. Disturbances were also included in Elliot and Murayama's MI CT-CM, but reasons for their inclusion was not presented. As such, unconstrained, constrained, and constrained with disturbances (disturbed) MI CT-CM models were tested. An unconstrained with disturbances MI CT-CM model was not tested because it is nonidentified (Kline, 2011).

The unconstrained, constrained, and disturbed MI CT-CM model were each fitted to the AGQ-R, drafted, and AGQR+ covariance matrices using maximum likelihood (ML) estimation method in AMOS 19, meaning that nine analyses were run. However, results for only five analyses were obtained. Programs employing ML calculate an initial solution to the model and perform subsequent calculations to lower the statistical criterion below a predetermined value (Kline, 2011). When the statistical criterion decreases below that value, minimum is said to have been achieved and AMOS will calculate fit indices as well as estimate model parameters. Minimum was achieved in all three models for the AGQ-R dataset and only the disturbed model for the other two datasets. As a result, the only models for which there are results are the 1) unconstrained with the AGQ-R data, 2) constrained with the AGQ-R data, 3) disturbed with the AGQ-R data, 4) disturbed with the drafted data, and 5) disturbed with the AGQR+ data.

Absolute, incremental, and parsimony-adjusted fit indices were evaluated for models where minimum was achieved. Absolute fit indices assess models as whole and can be interpreted as proportions of the covariance explained by the model (Kline, 2011). The absolute fit indices evaluated were χ^2 and the Goodness-of-Fit index (GFI). Incremental fit indices assess models in comparison to an independence model which assumes there is only one correct linear combination of the model population (Kline, 2011). The incremental fit index evaluated was the Normed Fit Index (NFI) and Comparative Fit Index (CFI). Parsimony-adjusted indices actually assess models in regards to their complexity; simpler models are favored over complex models (Kline, 2011). Hu and Bentler (1999) suggested fit indices of .95 or higher are evidence of good model fit. The parsimony-adjusted index evaluated was the root-mean-square-error-approximation (RMSEA). The RMSEA measures how poor a fit is, and an RMSEA = 0 suggests perfect fit. Brown and Cudeck (1993) advised that a heuristic of RMSEA less than .05 is evidence for a good model fit, but admitted this may not always be the case.

Other than model fit indices, the data were analyzed by looking at residual covariance and inspecting for any Heywood cases. Investigating residual covariance and checking for Heywood cases in addition to model fit indices is important because a model may show good fit but not be accurate. A residual covariance is the difference between the population covariance matrix and model covariance matrix (Kline, 2011). Positive residual covariance implies that the model underrepresented the relationship, and negative residual covariance implies that the model overrepresented the relationship. In an accurate model, most residual covariance should be less than an absolute value of 2. There is not a consensus on what "most" means, but the more residual covariance greater than the absolute value of 2 there are, the worse the model is (Kline, 2011). The residual covariance matrix was investigated as a whole to evaluate how well the model explains the data. Heywood cases are illogical and implausible parameter estimates which can occur when ML is used to test a model (Kline, 2011). Parameter estimates can include negative variances, correlation values greater than an absolute value of 1, and other extreme magnitudes. Heywood cases can occur because ML assumes a model is correct and will use any parameter values to achieve minimum. The presence of Heywood cases indicate that the model is not accurate, and the results were inspected for the presence of illogical and implausible parameter estimates.

Results

Descriptive statistics (i.e., mean, standard deviation, skew, and kurtosis) are provided after each drafted items in the appendices. Tanner (2011) suggested items with skew and kurtosis less than a magnitude of 1 behave with relative normality. As such, the hypothesized *performance-approach* items yielded the best normality because none of the items' skew or kurtosis exceeded a magnitude of 1. The hypothesized *performance-*

avoidance items yielded the worst normality because 11 items' skew and kurtosis exceeded a magnitude of 1.

The 2 x 2 Achievement Goal Factor Model

Presented in column 2 of Table 1 are values of fit indices for the unconstrained MI CT-CM model with the AGQ-R data. The model chi-square is statistically significant, and the exact-fit hypothesis is rejected because chi-squared (184.15) is larger than the number of degrees of freedom (49). This means there is a discrepancy between the population covariance matrix and those predicted by the model. The covariance matrix predicted by this model explains more than 90% of the total variability in the sample covariance matrix (GFI = .91). The relative fit of this model is approximately a 90% improvement (CFI = .90) over the independence model and an 87% (NFI = .87) improvement when adjusted for model complexity. The RMSEA for the 90% confidence interval is .11.

Table 1

Model Fit Indices

Index	Unconstrained	AGQ-R Constrained	Disturbed	drafted Disturbed	AGQR+ Disturbed
χ^2	184.149	237.563	177.195	3546.573	5596.155
df	49	53	49	1269	1947
p	.000	.000	.000	.000	.000
GFI	.907	.884	.913	.662	.588
CFI	.901	.864	.906	.722	.649
NFI	.871	.833	.876	.627	.549
RMSEA	.111	.122	.109	.081	.082

After inspecting the rest of the results for this model, a strong pattern was observed in the residual covariance matrix and no Heywood cases were found. As seen in Table 2, most of the standardized residual covariances in the unconstrained model with the AGQ-R data were within an acceptable range. However, item 11 on the AGQ-R has residual covariance that exceeds two for more than half of its correlations. The model may have underestimated the relation between item 11 and the *performance-avoidance* scale as well as the relation between item 11 and the *mastery-approach* scale.

Table 2

Standardized Residual Covariance Matrix for Unconstrained Model

	Performance-avoidance			Performance-approach			Mastery-avoidance			Mastery-approach		
	12	10	6	8	4	2	11	9	5	7	3	1
12	-0.33											
10	-0.29	-0.32										
6	-0.75	0.02	-0.28									
8	1.48	-0.59	-0.84	0.28								
4	1.12	-0.03	-0.10	0.66	0.34							
2	-0.46	-0.20	-1.70	0.54	0.11	0.25						
11	3.53	4.70	2.78	1.26	0.91	2.02	0.01					
9	-1.82	-0.60	0.15	-1.56	-0.32	-0.66	-0.61	0.02				
5	-2.01	-1.69	0.99	-1.67	-0.98	-0.82	-0.96	0.30	0.02			
7	1.68	1.61	0.64	0.04	-0.10	3.12	2.63	-0.44	-0.68	0.00		
3	0.30	0.42	-0.35	-2.86	-1.27	0.37	3.59	1.18	-0.74	1.01	0.00	
1	2.42	2.26	0.83	0.58	1.15	2.97	2.67	0.47	0.13	-0.43	-0.06	-0.01

Note. Numbers reflects item numbers in the AGQ-R.

Presented in column 3 of Table 1 are values of fit indices for the constrained MI CT-CM model with the AGQ-R data. The model chi-square is statistically significant, and that means there is a discrepancy between the population covariance and those predicted by the model. The covariance matrix predicted by this model explains approximately 88% of the total variability in the sample covariance matrix (GFI = .88). The relative fit of this model is approximately an 86% improvement (CFI = .86) over the independence model and an 83% (NFI = .83) improvement when adjusted for model complexity. The RMSEA for the 90% confidence interval is .12.

After inspecting the rest of the results for this model, no strong patterns were observed in the residual covariance matrix, nor were any Heywood cases found. There were residual covariances that exceeded a magnitude of 2, but there were few instances of this. As can be seen in Table 3, a relation between item 2 and *mastery-approach* may be underestimated.

Table 3

Standardized Residual Covariance Matrix for Constrained Model

	Performance-avoidance			Performance-approach			Mastery-avoidance			Mastery-approach		
	12	10	6	8	4	2	11	9	5	7	3	1
12	-0.01											
10	0.11	-0.01										
6	-0.68	-0.11	-0.42									
8	1.61	-0.57	-1.02	-0.70								
4	1.80	0.45	0.12	-0.45	0.00							
2	0.16	0.27	-1.46	-0.40	-0.34	0.00						
11	0.82	1.86	-0.03	0.45	0.16	1.36	0.00					
9	-1.15	0.03	0.60	-1.01	0.43	-0.02	-0.48	1.21				
5	-2.10	-1.89	0.63	-1.38	-0.58	-0.47	-1.65	6.27	0.00			
7	0.58	0.47	-0.50	-0.64	-0.55	2.70	0.20	-1.27	-1.88	0.00		
3	-0.81	-0.73	-1.48	-3.43	-1.72	-0.03	1.12	0.40	-1.95	0.91	0.00	
1	1.21	1.01	-0.44	-0.02	0.86	2.75	-0.10	-0.30	-1.15	-0.10	0.23	0.28

Note. Numbers reflects item numbers in the AGQ-R.

Presented in column 4 of Table 1 are values of fit indices for the disturbed MI CT-CM model with the AGQ-R data. The model chi-square is statistically significant, and that means there is a discrepancy between the population covariance and those predicted by the model. The covariance matrix predicted by this model explains more than 91% of the total variability in the sample covariance matrix (GFI = .91). The relative fit of this model is approximately a 91% improvement (CFI = .91) over the independence model and an 88% (NFI = .88) improvement when adjusted for model complexity. The RMSEA for the 90% confidence interval is .11.

After inspecting the rest of the results for this model, the same pattern in the residual covariance matrix of the unconstrained model was found along with Heywood cases. As can be seen in Table 4, this model may have underestimated the relation between item 11 and the *performance-avoidance* scale as well as the relation between item 11 and the *mastery-approach* scale. Concerning the Heywood cases, variance for the *performance-approach* and *performance-avoidance* disturbance terms was negative. This is in addition to the squared correlations for *performance-avoidance* and *performance-approach* exceeding 1.

Table 4

Standardized Residual Covariance Matrix for Disturbed Model

	Performance-avoidance			Performance-approach			Mastery-avoidance			Mastery-approach		
	12	10	6	8	4	2	11	9	5	7	3	1
12	-0.05											
10	-0.05	-0.05										
6	-0.42	0.36	-0.01									
8	1.37	-0.69	-0.86	-0.01								
4	1.06	-0.10	-0.08	0.19	0.03							
2	-0.50	-0.25	-1.68	0.13	-0.27	0.02						
11	3.88	5.06	3.12	1.50	1.19	2.25	-0.01					
9	-1.35	-0.12	0.67	-1.03	0.27	-0.15	-0.58	-0.04				
5	-1.55	-1.22	1.52	-1.17	-0.43	-0.34	-0.88	0.17	-0.02			
7	0.51	0.45	-0.43	-0.48	-0.63	2.61	2.48	-0.86	-1.02	0.00		
3	-0.86	-0.73	-1.41	-3.30	-1.76	-0.08	3.44	0.77	-1.07	1.08	0.00	
1	0.95	0.82	-0.50	-0.10	0.44	2.32	2.47	-0.07	-0.32	-0.44	-0.03	0.02

Note. Numbers reflects item numbers in the AGQ-R.

Presented in column 5 of Table 1 are values of fit indices for the disturbed MI CT-CM model with the drafted data. The model chi-square is statistically significant, and that means there is a discrepancy between the population covariance and those predicted by the model. The covariance matrix predicted by this model explains approximately 66% of the total variability in the sample covariance matrix (GFI = .66). The relative fit of the model is approximately a 72% improvement (CFI = .72) over the independence model and a 63% (NFI = .63) improvement when adjusted for model complexity. The RMSEA for the 90% confidence interval is .08.

After inspecting the rest of the results for this model, several strong patterns were found in the residual covariance matrix. A relation between the first drafted *performance-avoidance* item and the drafted *performance-approach* scale as well as the drafted *mastery-approach* scale may have been underestimated. Additionally, a relation between the tenth drafted *performance-avoidance* item and the drafted *performance-approach* scale as well as the drafted *mastery-approach* scale may have been underestimated. Numerous other residual covariances exceeded a magnitude of 2, but no other strong patterns were seen.

Heywood cases were also found in the results of this model. There was zero variance for the *mastery-avoidance* and *performance-avoidance* disturbance terms. The correlation between avoidance and *mastery-avoidance* as well as the correlation between performance and *performance-avoidance* exceeded 1. Additionally, the squared correlations for *mastery-avoidance* and *performance-avoidance* also exceeded 1.

Presented in column 6 of Table 1 are values of fit indices for the disturbed MI CT-CM model with the AGQR+ data. The model chi-square is statistically significant, and this means there is a discrepancy between the population covariance and those predicted by the model. The covariance matrix predicted by this model explains approximately 59% of the total variability in the sample covariance matrix (GFI = .59). The relative fit of this model is approximately a 65% improvement (CFI = .65) over the independence model and a 55% (NFI = .55) improvement when adjusted for model complexity. The RMSEA for the 90% confidence interval is .08.

After inspecting the results for this model, all the patterns found in previous residual covariance matrices reappeared. Several new patterns arose as well. The relation between the *performance-avoidance* scale and the rest of the AGQ-R as well as the drafted *performance-approach* items may have been underestimated. The relation between item 8 on the AGQ-R and the drafted *performance-avoidance* items was also underestimated. Numerous other covariances scattered throughout the matrix also exceeded a magnitude of 2, but no other strong patterns were found.

The Heywood cases found in the results for this model were the same as the ones observed in the disturbed model with drafted data. Zero variance for the *mastery-avoidance* and *performance-avoidance* disturbance terms was observed. The correlation between avoidance and *mastery-avoidance* as well as the correlation between performance and *performance-avoidance* exceeded one. The squared correlations for *mastery-avoidance* and *performance-avoidance* also exceeded 1.

Coefficient Alpha and Variance

Coefficient alpha is a measure of reliability and calculates the interrelatedness of items based on variance (Crocker & Algina, 2008). Coefficient alpha was calculated for the *mastery-approach, mastery-avoidance, performance-approach,* and *performance-avoidance* scales from the AGQ-R and AGQ-R+ datasets. To test whether there were significant differences between the coefficient alphas and variances from the AGQ-R and AGQ-R+ datasets, the AGQ-R+ dataset was resampled using the bootstrap method (Efron, 1979).

Bootstrapping is a statistical resampling technique that was necessary to determine whether there was a statistically significant difference between the two coefficient alphas. Bootstrapping generates additional datasets by randomly sampling with replacement cases from the original dataset (Wilson & Batterham, 1999; Zhu, 1997). This means that a new dataset was created by randomly selecting participants' responses from the original dataset until the new dataset had the same number of cases as the original dataset (i.e., 300 in this study). The bootstrap typically generates 1,000 new datasets, which was the case in this research (Mooney & Duval, 1993). Coefficient alpha and variance were calculated for each bootstrapped dataset and allowed for significant differences to be tested by searching for overlapping values.

There was a statistically significant difference between the reliabilities of the scores for the *mastery-approach*, *performance-approach*, and *performance-avoidance* scales on the AGQ-R and the AGQ-R+ using bootstrapping. However, only the reliabilities of the scores for the *mastery-approach* and *performance-approach* were in the hypothesized direction (i.e., the reliabilities of the scores from the AGQ-R+ were greater than the reliabilities of the scores from the AGQ-R). The reliability of the *performance-avoidance* scores on the AGQ-R was significantly higher than the reliability of the *performance-avoidance* scores from the AGQ-R+. There was no significant difference between the *mastery-avoidance* scores. These results are presented in Table 5.

The variances of the resampled scores from each subscale on the AGQ-R and AGQ-R+ datasets are also presented in Table 5. Variance decreased in each subscale. All the differences, except for the *performance-approach*, were significant.

Table 5

Reliability and Variance

	Coefficient Alpha			Variance		
	AGQ-R	AGQR+	*p*	AGQ-R	AGQR+	*p*
Mastery-Approach	.75	.91	.000	0.41	0.28	.000
Mastery-Avoidance	.69	.66	.234	1.11	0.16	.000
Performance-Approach	.78	.93	.000	0.63	0.60	.234
Performance-Avoidance	.80	.72	.005	0.99	0.15	.000

Note: Statistics obtained from bootstrapping (*N* = 1000) and a two-tailed distribution.

Discussion

Hypotheses

The first research question asked how the AGQ-R with drafted items would load on to the 2 x 2 factor model of achievement goal theory. It was hypothesized that the drafted data set would fit the 2 x 2 factor model better than the AGQ-R+ data set and that the AGQ-R data set would fit the model better than the AGQ-R+ data set. These two hypotheses were supported. However, all of the models explained less than 95% of the total variability in the sample. The relative fit for each model is less than a 95% improvement over the independence model, where no variables are related. These results indicate that the MI CT-CM models are a poor fit and that the AGQ-R, drafted, as well as the AGQ-R+ data do not fit the 2 x 2 factor model of achievement goal theory.

The large magnitudes in the residual covariance matrices may explain why the model fit indices were not as good as the ones obtained by Elliot and Murayama (2008) for their MI CT-CM model. In every model, except for the constrained one, the covariance of item 11 on the AGQ-R with most every other item was underestimated by the model. Item 11 is a *mastery-avoidance* measure, but it may be related to *mastery-approach* and *performance avoidance* in a way not hypothesized by the model.

According to the 2 x 2 factor structure, item 11 should only be related to *mastery-approach* goals through a *mastery* competency. The MI CT-CM model assumes this, but the model does not fit the data collected here. Face validity indicates that item 11 does not belong with the *mastery-approach* measures and a scale analysis confirms this. As such, item 11 may relate to *mastery-approach* goals in an indirect way simply not predicted by the model. However, this pattern is not exhibited by the other *mastery-avoidance* items on the

AGQ-R, which may mean that there is simply an issue with the item and not an issue with the relation between *mastery-avoidance* and *mastery-approach* goals.

Additionally, this pattern may be an indication of a spurious relation resulting from the sample and not the 2 x 2 model structure. Elliot and Murayama (2008) also used a MI CT-CM model, but they did not provide or discuss residual covariance. As such, it is difficult to determine whether this relation is a result of the sample used, the model structure, or a combination of the two. Further replication is required to find evidence of the cause. Also, this same logic can be applied to the relation between item 11 and *performance-avoidance* measures.

The residual covariance matrices are important to review in modeling because they highlight relationships between each variable as well as between variables and the model itself. For example, the residual covariance matrix for the disturbed model showed that model has problems (i.e., it inaccurately represents relationships between items) even though it has the best model fit. Additionally, the residual covariance matrices indicate that the drafted items as a whole do not function well in the 2 x 2 achievement goal theory model. The findings for the drafted items in each subscale (i.e., each achievement goal) are mixed, however, and discussed next.

The hypothesis for the second research question, that reliability would increase by adding the drafted items, was supported for the *mastery-approach* and *performance-approach* scales but was not confirmed for the *mastery-avoidance* and *performance-avoidance* scales. Reliability increased for *mastery-approach* and *performance-approach* scores because the number of items in the scale increased and because the added items also measure the same construct. However, several of the hypothesized *mastery-approach* items exceeded normality though (i.e., kurtosis exceeded a magnitude of 1), but none of the hypothesized *performance-approach* items did. The non-normal *mastery-approach* items should be modified in future use because their peakedness indicates they did not differentiate between students.

Similarly, the scores for numerous drafted items from *mastery-avoidance* and *performance-avoidance* subscales did not follow a normal distribution. The hypothesis for the second research question was not supported in these cases. Reliability decreased for the scores from these items because the added items did not measure the same construct. It is important to note that the decrease in these reliabilities reflect the responses to all the items in the scale. This means that researchers should not include all of the drafted items in the future, but they can include certain drafted items. Even though the reliability significantly decreased after all of the hypothesized *mastery-avoidance* items were added, researchers may want to consider adding items whose skew as well as kurtosis fell within acceptable limits and that have face validity. For example, the skew and kurtosis for item 15 (*I do not want to forget the material I learn in class)* are within ±1, which means researchers may want to use that item when measuring *mastery-avoidance*.

Additionally, the fact that the reliability of the responses to the *approach* scales increased while the reliability of the responses to the *avoidance* scales decreased highlights a limitation of this study. Social desirability was not controlled for. Identifying strongly with the *mastery-approach* and *performance-approach* measures may be very appealing to university students. However, the reliabilities for responses to both of those scales were less than .8. If social desirability had been a threat, then it seems like the reliabilities for the responses to the AGQ-R should have been higher. It is also possible that the hypothesized *approach* items were simply written in a way that measured *mastery-approach* and *performance-approach* items better than how the hypothesized *avoidance* items were written.

The hypothesis for the third research question, that variance would increase by adding drafted items, was not supported in any of the sub-scales. The data actually support the opposite of the hypothesis (i.e., variance decreased after the drafted items were added) and was significant for the *mastery-approach, mastery-avoidance,* and *performance-avoidance* sub-scales.

Muis et al. (2009) suggested the addition of cleverly worded items to the AGQ could increase the variance of responses. As such, the decrease in variance following the addition of the drafted items may be indicative of the drafted items not being worded in such a way that differentiated students. Since the drafted items included particular academic behaviors, the items may have become redundant for students. Academic behavior may be hierarchical. In other words, students need to complete one task before they can complete another. For example, a student may need to write down notes during lecture before he/she can study. This means if the student does not write down notes, then the student cannot study.

Limitations and Future Research

This study added to the research in achievement motivation because it replicated Elliot and Murayama's (2008) MI CT-CM factor model of achievement goal theory as well as increased the number of items available to measure *mastery-approach* and *performance-approach* goals. However, there were limitations to this study. A sample of convenience was used to collect the data, and this affects the generalizability of the findings. This means that adding the drafted items to the *mastery-approach* scale may not necessarily increase the reliability of the responses as was the case in this study.

Another limitation was the method in which the survey was administered. Of the 300 surveys completed, more than 200 were completed online. This means that the researcher was not able to control for the environment which threatened the reliability. This can be addressed in future research by having participants complete the survey in the same setting at the same time under controlled conditions.

Furthermore, social desirability was not controlled. Participants may not have answered truthfully, and this may have affected the results. This can be addressed in the future by including social desirability scales along with the survey.

The 2 x 2 factor model of achievement goal theory is commonly used in the field of achievement motivation and that is why future research should continue to replicate the MI CT-CM model until it can be statistically confirmed. Determining whether constraining second order factors is theoretically reasonable would also bolster research quality. Continuing to investigate this model can help researchers learn not only how goals are related to other academic constructs, but how goals are related to one another as well.

Future research should also try to build upon the knowledge of *mastery-approach* goals and extend it to *performance-approach, mastery-avoidance,* and *performance-avoidance* goals as well. Scales with high reliability and high variance, like the one created by adding the drafted items to the *mastery-approach* scale, can allow individual differences to be measured in the manner Muis et al. (2009) described. Adding to the knowledge of achievement goals does not simply mean analyzing more items. Adding to the knowledge of achievement goals means performing qualitative research like Urdan and Mestas (2006) so that nomological changes can be made to items and the model. Then items and models can be analyzed to determine whether researchers have an accurate model, which should better than the models in this research.

Including all of the drafted items in MI CT-CM model may very likely contributed to the poor fit, but was necessary in some ways since there were no previous data to make a-priori assumptions. Future research can also include strategic use of the drafted items in the 2 x 2 factor model now that there are data on them. For example, most of the hypothesized *performance-approach* items had acceptable skew and kurtosis, which means it might be reasonable to use some of those items when measuring achievement goals. Additionally, items with unacceptable skew and/or kurtosis can be reworded and used as well when measuring achievement goals. Adding items and collecting data on them on in a process like this adds to the literature and research of achievement goal theory.

REFERENCES

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*, 261-271. doi:10.1037/0022-0663.84.3.261.

Atkinson, J. W. (1983). *Personality, motivation, and action: Selected papers*. New York, NY: Praeger.

Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Church, M., Elliot, A., & Gable, S. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology*, *93*, 43-54. doi:10.1037/0022-0663.93.1.43.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory.* Mason, OH: Cengage Learning

Eccles, J. (1983). Expectancies, values, and academic behavior. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75-146). San Francisco, CA: W.H. Freeman.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait–multimethod models: A multiple-indicator CT-C (M-1) model. *Psychological Methods*, *8*, 38–60.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1 – 26.

Elliot, A. (1997). Integrating the “classic” and “contemporary” approaches to achievement motivation: A hierarchical model of approach and avoidance *achievement* motivation.

In M. Maehr & P. Pintrich (Eds.), *Advances in motivation and achievement Vol. 10,* (pp. 243–279). Greenwich, CT: JAI Press.

Elliot, A. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, *34*, 149-169.

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, *70,* 461-475.

Elliot, A., & McGregor, H. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *76*, 628-644. doi:10.1037/0022-3514.76.4.628.

Elliot, A., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, *100*, 613-628. doi:10.1037/0022-0663.100.3.613.

Finney, S. J., Pieper, S. L., & Barron, K. E. (2004). Examining the psychometric properties of the Achievement Goal Questionnaire in a general academic context. *Educational and Psychological Measurement*, *64*, 365-382. *doi:10.1177/0013164403258465*

Fulmer, S., & Frijters, J. (2009). A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review*, *21*, 219-246. doi:10.1007/s10648-009-9107-x.

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

Hulleman, C., Schrager, S., Bodmann, S., & Harackiewicz, J. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels?. *Psychological Bulletin*, *136*, 422-449. doi:10.1037/a0018947.

Jackson, D. N., Ahmed, S. A., & Heapy, N.A. (1976). Is achievement a unitary construct?, *Journal of Research in Personality*, 10, 1. doi:10.1016/0092-6566(76)90079-9

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford.

Levin, B. (2007). Schools, poverty, and the achievement gap. *Phi Delta Kappan*, *89*, 75-76.

Linnenbrink, E. (2005). The dilemma of performance-approach goals: The use of multiple goal contexts to promote students' motivation and learning. *Journal of Educational Psychology*, *97*, 197-213. doi:10.1037/0022-0663.97.2.197.

Maloney, P., Grawitch, M. J., & Barber, L. K. (2011). Strategic item selection to reduce survey length: Reduction in validity?. *Consulting Psychology Journal: Practice And Research*, *63*, 162-175. doi:10.1037/a0025604

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait–multimethod analyses: Application of second order confirmatory factor analysis. *Journal of Applied Psychology*, *73*, 107–117.

McCaslin, M. (2006). Student motivational dynamics in the era of school reform. *Elementary School Journal*, *106*, 479-490.

Middleton, M., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology*, *89*, 710-718. doi:10.1037/0022-0663.89.4.710.

Midgley, C., Kaplan, A., & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost?. *Journal of Educational Psychology, 93,* 77-86. doi:10.1037/0022-0663.93.1.77.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., … Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS),* Ann Arbor, MI: University of Michigan.

Mooney, C. Z. & Duval, R. D. (1993). Bootstrapping: A nonparametric approach to statistical inference. Newbury Park, CA: Sage.

Muis, K. R., Winne, P. H., & Edwards, O. V. (2009). Modern psychometrics for assessing achievement goal orientation: A Rasch analysis. *British Journal of Educational Psychology*, 79, 547-576. doi:10.1348/000709908X383472

Mulvenon, S., Wang, K., McKenzie, S., & Airola, D. (2006). A case study: Using geographic information systems for education policy analysis. *Educational Research Quarterly*, *30*, 44-54.

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2008) http://www2.ed.gov/policy/elsec/leg/esea02/index.html.

Pintrich, P. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*, 667-686. doi:10.1037/0022-0663.95.4.667.

Pintrich, P., & de Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33-40. doi:10.1037/0022-0663.82.1.33.

Raynor, J. O. (1969). Future orientation and motivation of immediate activity: An elaboration of the theory of achievement motivation. *Psychological Review*, 76, 606-610. doi:10.1037/h0028291

Tanner, D. (2011). *Using statistics to make educational decisions*. Thousand Oaks, CA: Sage Publications, Inc.

Urdan, T., & Mestas, M. (2006). The goals behind performance goals. *Journal of Educational Psychology*, *98*, 354-365. doi:10.1037/0022-0663.98.2.354.

Wilson, K. K., & Batterham, A. A. (1999). Stability of questionnaire items in sport and exercise psychology: Bootstrap limits of agreement. *Journal of Sports Science*, *17*, 725-734.

Wolters, C. A., Yu, S. L., & Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences*, *8*, 211-238. doi:10.1016/S1041-6080(96)90015-1

Zhu, W. (1997). Making bootstrap statistical inferences: A tutorial. *Research Quarterly for Exercise and Sport*, *68*, 44 – 55.

APPENDIX A

SUMMARY OF CHANGES MADE TO THE AGQ

Mastery-approach items

1. My aim is to completely master the material presented in class (originally: I desire to completely master the material presented in class).[a]

3. My goal is to learn as much as possible (originally: I want to learn as much as possible from this class).[a]

7. I am striving to understand the content of this course as thoroughly as possible (originally: It is important for me to understand the content of this course as thoroughly as possible).[a]

Mastery-avoidance items

5. My aim is to avoid learning less than I possibly could (originally: I worry that I may not learn all that I possibly could in this class).[a]

9. My goal is to avoid learning less that it is possible to learn (originally: I am often concerned that I may not learn all that there is to learn in this class).[a]

11. I am striving to avoid an incomplete understanding of the course material (originally: Sometimes I'm afraid that I may not understand the content of this class as thoroughly as I would like).[a, b]

Performance-approach items

2. I am striving to do well compared to other students (originally: It is important for me to do well compared to others in this class).[a]

4. My aim is to perform well compared to other students (originally: My goal in this class is to get a better grade than most of the other students).[c]

8. My goal is to perform better than other students (originally: It is important for me to do better than other students).[a]

Performance-avoidance items

6. My goal is to avoid performing poorly compared to others (originally: My goal in this class is to avoid performing poorly).[d]

10. I am striving to avoid performing worse than others (originally: My fear of performing poorly in this class is often what motivates me).[a, e]

12. My aim is to avoid doing worse than other students (originally: I just want to avoid doing poorly in this class).[a, f]

Note: This is adapted from “On the Measurement of Achievement Goals: Critique, Illustration, and Application,” by A. J. Elliot and K. Murayama, 2008, *Journal of Educational Psychology*, *100*, 613-628. Copyright 2008 by the American Psychological Association.

[a] The word “aim”, “goal”, or “striving” was added to the item. [b] Affective aspect was removed. [c] References to grades and the word “most” were removed. [d] Reference to other students was added. [e] Reason for the goal was removed. [f] The word “just” was removed.

APPENDIX B

HYPOTHESIZED MASTERY-APPROACH ITEMS

Hypothesized *Mastery-Approach* Items ($\alpha = .75$) with means (M), standard deviations (SD), skew (g_1), and kurtosis (g_2).

13. I want to learn as much of the material taught in class as I possibly can ($M = 4.31$, $SD = 0.74$, $g_1 = -0.96$, $g_2 = 1.05$)
18. I study to learn more about the material being taught ($M = 3.86$, $SD = 0.88$, $g_1 = -0.26$, $g_2 = -0.61$).
24. When I receive graded assignments back, I look for feedback and comments in order to learn more about the material ($M = 4.00$, $SD = 0.90$, $g_1 = -.49$, $g_2 = -0.55$).
25. I pay attention in class so that I can learn the material ($M = 4.38$, $SD = 0.71$, $g_1 = -1.03$, $g_2 = 1.20$).
32. I work on assignments because I want to learn the material being taught ($M = 4.19$, $SD = 0.81$, $g_1 = -0.86$, $g_2 = 0.70$).
38. I think about what was said in class because it helps me learn the material ($M = 4.31$, $SD = 0.76$, $g_1 = -0.76$, $g_2 = -0.25$).
42. I put effort into my work because I want to learn the material being taught ($M = 4.34$, $SD = 0.76$, $g_1 = -0.98$, $g_2 = 0.75$).
45. I answer problems I am not sure how to do because I have to try in order to learn the material ($M = 4.00$, $SD = 0.88$, $g_1 = -0.86$, $g_2 = 0.94$).
46. I ask my peers or teacher for help so I can learn the material ($M = 3.72$, $SD = 1.02$, $g_1 = -0.47$, $g_2 = -0.28$).
51. I take notes because it helps me learn the material ($M = 4.58$, $SD = 0.68$, $g_1 = -1.80$, $g_2 = 3.76$).

53. I review class notes because it helps me learn the material (*M* = 4.46, *SD* = 0.73, g_1 = -1.28, g_2 = 1.44).

58. I use what I learn because it helps me understand the material (*M* = 4.20, *SD* = 0.80, g_1 = -0.74, g_2 = 0.15).

63. I talk about what was said in class because it helps me learn the material (*M* = 3.89, *SD* = 1.00, g_1 = -0.78, g_2 = 0.40).

APPENDIX C

HYPOTHESIZED MASTERY-AVOIDANCE ITEMS

Hypothesized *Mastery-Avoidance* Items (α = .75) with means (*M*), standard deviations (*SD*), skew (g_1), and kurtosis (g_2).

15. I do not want to forget the material I learn in class (M = 4.21, SD = 0.87, g_1 = -0.88, g_2 = 0.41).
17. I put enough effort into my work so that I do not forget the material being taught (M = 3.80, SD = 0.80, g_1 = -0.28, g_2 = -0.04).
20. I do not try to answer a problem if I forgot the material (M = 2.20, SD = 1.06, g_1 = 0.53, g_2 = -0.48).
27. I do not take notes because it does not help me remember the material (M = 1.46, SD = 0.83, g_1 = 1.20, g_2 = 3.66).
29. I do not review the notes because it does not help me remember the material (M = 1.45, SD = 0.77, g_1 = 1.89, g_2 = 3.61).
34. I do not ask for help because I know I can remember the material (M = 2.29, SD = 1.05, g_1 = 0.37, g_2 = -0.55).
40. I do not talk about what was said in class because it does not help me remember the material (M = 1.69, SD = 0.86, g_1 = 1.12, g_2 = 0.79).
43. I do not study because I remember the material that was taught (M = 1.76, SD = 0.90, g_1 = 0.95, g_2 = 0.05).
47. I do not use what I learn because it does not help me remember material (M = 1.65, SD = 0.85, g_1 = 1.38, g_2 = 1.84).
49. I do not pay attention in class because it does not help me to remember the material (M = 1.44, SD = 0.76, g_1 = 1.82, g_2 = 2.75).

56. I do not work on assignments because I forgot the material (*M* = 1.48, *SD* = 0.70, g_1 = 1.43, g_2 = 1.67).

59. When I receive a graded assignment back, I do not look at it because it will not help me remember the material (*M* = 1.52, *SD* = 0.88, g_1 = 1.92, g_2 = 3.57).

61. I do not think about what was said in class because it does not help me remember the material (*M* = 1.54, *SD* = 0.83, g_1 = 1.71, g_2 = 2.92).

APPENDIX D

HYPOTHESIZED PERFORMANCE-APPROACH ITEMS

Hypothesized *Performance-Approach* Items ($\alpha = .92$) with means (M), standard deviations (SD), skew (g_1), and kurtosis (g_2).

14. I want to do better than the other students in class (M = 4.15, SD = 0.89, g_1 = -0.97, g_2 = 0.73).
22. I ask my teacher, or another authority, for help so I can do better than other students (M = 3.16, SD = 1.06, g_1 = -0.11, g_2 = -0.29).
23. I use what I learn to demonstrate that my abilities are better than the other students' abilities (M = 3.21, SD = 1.16, g_1 = -0.20, g_2 = -0.54).
26. I think about what was said in class so that I can perform better than other students (M = 3.75, SD = 1.00, g_1 = -0.64, g_2 = 0.22).
31. I put more effort into my work to show that I am better than other students (M = 3.23, SD = 1.20, g_1 = -0.29, g_2 = -0.67).
33. I answer problems I am not sure how to do because I cannot do better than other students if I do not try (M = 3.46, SD = 1.15, g_1 = -0.42, g_2 = -0.41).
37. I pay attention in class in so that I can perform better than other students (M = 3.89, SD = 1.02, g_1 = -0.84, g_2 = 0.38).
41. I review class notes because I want to do better than the other students (M = 3.84, SD = 1.08, g_1 = -0.77, g_2 = 0.04).
44. I work on assignments because I want to show how much better my performance is compared to other students' performance (M = 3.36, SD = 1.24, g_1 = -0.33, g_2 = -0.87).
48. When I receive graded assignments back, I look at feedback if I know other students have done better than me (M = 3.59, SD = 1.19, g_1 = -0.52, g_2 = -0.54).

52. I talk about what was said in class so that I can perform better than other students (M = 3.49, SD = 1.14, g_1 = -0.36, g_2 = -0.54).
55. I study so that I can show that I am better than other students (M = 3.32, SD = 1.31, g_1 = -0.35, g_2 = -0.89).
62. I take notes so that I can do better than other students (M = 3.80, SD = 1.15, g_1 = -0.83, g_2 = -0.02).

APPENDIX E

HYPOTHESIZED PERFORMANCE-AVOIDANCE ITEMS

Hypothesized *Performance-Avoidance* Items (α = .77) with means (*M*), standard deviations (*SD*), skew (g_1), and kurtosis (g_2).

16. I do not want to do worse than the other students in class (*M* = 4.38, *SD* = 0.84, g_1 = -1.70, g_2 = 3.49).
19. I do not work on assignments because other students do not work on assignments (*M* = 1.63, *SD* = 0.96, g_1 = 1.64, g_2 = 2.22).
21. I do not answer a problem on an assignment if I think other students will not answer it also (*M* = 1.54, *SD* = 0.82, g_1 = 1.47, g_2 = 1.51).
28. I do not talk about what was said in class because other students do not (*M* = 1.87, *SD* = 0.93, g_1 = 0.62, g_2 = -0.17).
30. I do not study because I know other students do not study (*M* = 1.21, *SD* = 0.54, g_1 = 3.29, g_2 = 13.79).
35. I do not use what I learn because I do not see other students using what they learn (*M* = 1.56, *SD* = 0.79, g_1 = 1.38, g_2 = 1.56).
36. When I receive a graded assignment back, I do not look at it because it will not help me do as well as other students (*M* = 1.42, *SD* = 0.82, g_1 = 2.15, g_2 = 4.44).
39. I do not take notes because other students do not take notes (*M* = 1.28, *SD* = 0.58, g_1 = 2.13, g_2 = 4.37).
50. I do not think about what was said in class because other students do not (*M* = 1.41, *SD* = 0.74, g_1 = 2.08, g_2 = 4.71).
54. I put enough effort into my work so that I will not do worse than other students (*M* = 3.89, *SD* = 1.03, g_1 = -0.95, g_2 = 0.69).

57. I do not ask for help because I think other students will think that I am not as good as they are (M = 1.33, SD = 0.64, g_1 = 2.07, g_2 = 4.01).

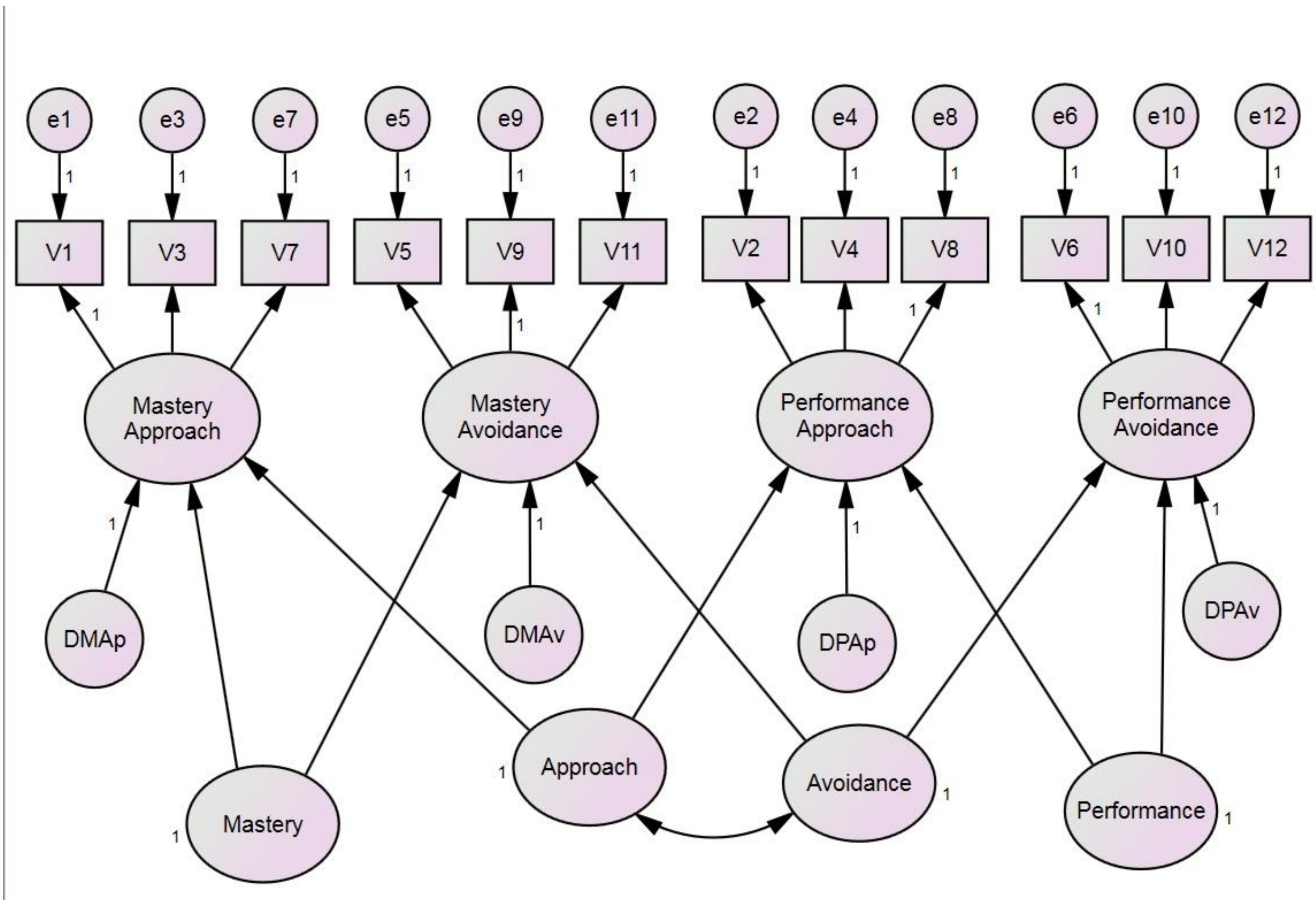
60. I do not pay attention in class because other students do not pay attention (M = 1.32, SD = 0.65, g_1 = 2.45, g_2 = 6.95).

64. I do not review class notes because other students do not review the notes (M = 3.80, SD = 1.15, g_1 = -0.83, g_2 = -0.02).

APPENDIX F

THE DISTURBED MI CT-CM 2 X 2 FACTOR MODEL

The Disturbed MI CT-CM 2 x 2 Factor Model



Note: Variables correspond to item numbers on the AGQ-R (e.g., V1 is item 1).