Test emotions, value, and self-efficacy: A longitudinal model predicting examinee effort and performance on a low-stakes test

Paulius Satkus
Test emotions, value, and self-efficacy: A longitudinal model predicting examinee effort and performance on a low-stakes test

Paulius Satkus

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

May 2019

FACULTY COMMITTEE:

Committee Chair: Sara J. Finney

Committee Members/ Readers:

Dena A. Pastor

John D. Hathcoat
Acknowledgements

First, I would like to thank my advisor and mentor, Dr. Sara Finney. You are the most incredible professor I have ever met. Your high standards and timely, candid feedback has been invaluable during this project. I could not envision a better mentor to learn from at the beginning of my professional career. You were always available and willing to help when I was lost. Your professional dedication is unmatched, and you are a constant inspiration to everyone around you. I honestly cannot express well enough how much you of an impact you have made on me throughout these two years.

Second, I want to say a huge thanks to my committee members, Drs. Dena Pastor and John Hathcoat. I appreciate the time and energy you spent on this project. Your suggestions and feedback have made me a better researcher.

Third, I want to say thanks to all the faculty and students at Center for Assessment and Research Studies. I always felt support and encouragement from all of you.

Fourth, thank you, Beth, my academic sibling throughout these two years. I appreciate your patience and willingness to hear me out.

Lastly, thank you, Olivia. This project could not have been possible without your support and encouragement. You were always there when I needed you. I am extremely grateful for you.
# Table of Contents

Acknowledgements .......................................................................................... ii  
List of Tables ........................................................................................................ v  
List of Figures .......................................................................................................... vi  
Abstract .................................................................................................................. vii  

Chapter 1: Introduction ......................................................................................... 1  
  Expectancy-Value Theory Applied to Examinee Effort .................................. 3  
  Demands-Capacity Model of Examinee Effort .................................................. 8  
  Control-Value Theory: Explicit Incorporation of Achievement Emotions ...... 9  
  Purpose of The Current Study ............................................................................. 13  

Chapter 2: Literature Review ............................................................................... 16  
  Achievement Emotions ....................................................................................... 16  
    Control-value theory of achievement emotions. ............................................. 18  
    Antecedents of emotions. ................................................................................. 20  
    Antecedents to control and value appraisals. ................................................. 30  
    Outcomes of emotions. .................................................................................... 32  
    Reciprocal causation in CV theory ................................................................. 34  
  Mediation ............................................................................................................ 41  
    Research design and inferences ..................................................................... 45  

Chapter 3: Method ................................................................................................. 60  
  Participants and Procedure ................................................................................ 60  
  Measures ............................................................................................................. 62  
    Quantitative and scientific reasoning ............................................................ 62  
    Test emotions. .................................................................................................. 64  
    Perceived test importance and examinee effort ......................................... 65  
    Self-efficacy ..................................................................................................... 66  
  Data Analysis ..................................................................................................... 67  
    Model-Data fit. .................................................................................................. 67  
    Autoregressive panel models ......................................................................... 69  
    Proposed autoregressive panel models. ......................................................... 70  
    Estimated autoregressive panel models. ....................................................... 77  

Chapter 4: Results ................................................................................................. 84  
  Descriptive Statistics ......................................................................................... 84  
  Stability .............................................................................................................. 85  
  Equilibrium ........................................................................................................ 87  
  Autoregressive panel models .......................................................................... 89  
    Research question 1. ....................................................................................... 89  
    Research question 2. ....................................................................................... 89  
    Research question 3. ....................................................................................... 91  
    Research question 4. ....................................................................................... 92  
    Research question 5. ....................................................................................... 94  
    Research question 6. ....................................................................................... 95  
    Research question 7. ....................................................................................... 95  
  Interpretation of Direct and Indirect Effects .................................................. 98  
    Direct effects. .................................................................................................. 100  
    Indirect Effects. ............................................................................................... 103
List of Tables

Table 1: Three-dimensional model of achievement emotions ........................................ 127
Table 2: Determinants and outcomes of achievement emotions .................................. 128
Table 3: Comparison between types of subjective value in Expectancy-Value and Control-Value theory ................................................................. 129
Table 4: Summary of research designs and the inferences that can be drawn from each ................................................................. 130
Table 5: Demographics of the sample ........................................................................... 131
Table 6: Descriptive statistics and reliability for gathered variables .............................. 132
Table 7: Proposed research questions, analyses, expected results, and the implications of results .............................................................................................. 134
Table 8: Actual research questions, analyses, expected results, and the implications of results .............................................................................................. 136
Table 9: Model summary for mean stability ................................................................. 138
Table 10: Bivariate correlations between the same constructs at different times ............ 139
Table 11: Model summary for evaluating equality of variances .................................. 140
Table 12: Model fit for each estimated panel model ...................................................... 141
Table 13: Direct effects of constructs at time 1 on constructs at time 2 from Adjacent Reciprocal Effects Model ................................................................. 142
Table 14: Direct effects of constructs at time 2 on constructs at time 3 from Adjacent Reciprocal Effects Model ........................................................................ 143
Table 15: Indirect effects of constructs at time 1 on constructs at time 3 from Adjacent Reciprocal Effects Model ................................................................. 144
Table 16: Direct effects of constructs at time 1 on constructs at time 2 from Complete Reciprocal Effects Model ................................................................. 145
Table 17: Direct effects of constructs at time 2 on constructs at time 3 from Complete Reciprocal Effects Model ........................................................................ 146
Table 18: Indirect effects of constructs at time 1 on constructs at time 3 from Complete Reciprocal Effects Model ................................................................. 147
Table 19: Summary of changes for the direct effects between the two reciprocal effects models .............................................................................................. 148
List of Figures

Figure 1: The Expectancy-Value (EV) theory of examinee effort .................................. 149
Figure 2: The Control-Value (CV) theory of achievement emotions ............................... 150
Figure 3: Depiction of simple regression (a), partial mediation (b), and complete mediation (c) .......................................................................................................................... 151
Figure 4: Popular research designs for estimating mediation: A) Cross-sectional, B) Sequential, C) Half-Longitudinal, D) Fully-Longitudinal ............................................. 152
Figure 5: A path model with five waves of measurement ................................................. 153
Figure 6: Proposed Autoregressive Model A .................................................................. 154
Figure 7: Proposed Autoregressive Model B ................................................................. 155
Figure 8: Proposed EV Theory Model A ......................................................................... 156
Figure 9: Proposed EV Theory Model B ................................................................. 157
Figure 10: Proposed Combined EV and CV Theory Model A ........................................ 158
Figure 11: Proposed Combined EV and CV Theory Model B ........................................ 159
Figure 12: Proposed CV Theory Model A ................................................................. 160
Figure 13: Proposed CV Theory Model B ................................................................. 161
Figure 14: Proposed Adjacent Reciprocal Effects Model A ........................................... 162
Figure 15: Proposed Adjacent Reciprocal Effects Model B ........................................... 163
Figure 16: Proposed Complete Reciprocal Effects Model A .......................................... 164
Figure 17: Proposed Complete Reciprocal Effects Model B .......................................... 165
Figure 18: Autoregressive Model A ............................................................................. 166
Figure 19: Autoregressive Model B ............................................................................. 167
Figure 20: EV Theory Model ..................................................................................... 168
Figure 21: Combined EV and CV Theory Model ......................................................... 169
Figure 22: CV Theory Model ..................................................................................... 170
Figure 23: Adjacent Reciprocal Effects Model ............................................................. 171
Figure 24: Complete Reciprocal Effects Model ............................................................ 172
Abstract

The validity of scores from low-stakes tests may be compromised by examinee motivation. Expectancy-Value theory (EV) has been used to frame the antecedents of examinee motivation in low-stakes testing contexts. According to EV theory, the perceived value of the test and the expectancy to succeed on the test directly affect examinee effort, which then affects test performance. Cross-sectional research studies in low-stakes testing contexts offer some support of EV theory. Control-Value theory (CV) serves as another theory to understand motivation toward a task. CV theory encompasses the constructs of expectancy and value from EV theory, but incorporates test emotions as mediators of the effects of expectancy and value on motivation. Unfortunately, the role of emotions when studying examinee motivation in low-stakes contexts has been largely ignored. The purpose of the current study was to examine the direct and indirect effects of perceived value and test emotions on examinee effort during a low-stakes test. To address the purpose of the study and the limitations of previous cross-sectional designs, several panel models were estimated using longitudinal data. Specifically, value, test emotions, and motivation were measured three times throughout the test. Two models based on CV theory fit the data well. In these models, when controlling for prior test emotions and prior test performance, the effects of perceived value on subsequent examinee effort were not statistically significant at any point during the test, whereas the effects of several test emotions on subsequent examinee effort were significant. The results suggest that practitioners may need to shift their attention to other constructs that may impact examinee effort and test performance. Empirical studies on test emotions are lacking, thus several areas of research avenues are proposed.
Chapter 1: Introduction

Numerous empirical studies have been conducted to examine the validity of test scores due to concerns of low examinee effort in low-stakes testing contexts (for a review see Wise & DeMars, 2005; Wise & Smith, 2011). Examinee effort is defined as “student’s engagement and expenditure of energy towards the goal of attaining the highest possible score on the test” (Wise & DeMars, 2005). Low-stakes tests, by definition, have minimal or no consequences for students who complete them.

Researchers have shown that students invest variable levels of effort when completing low-stakes tests (e.g., Sundre & Kitsantas, 2004; Thelk, Sundre, Horst, & Finney, 2009). The concern when interpreting test scores from low-stakes tests arises because variability in examinee effort is related to variability in test scores (e.g., Cole, Bergin, & Whitaker, 2008). Therefore, differences in test scores across students represent not only differences in students’ abilities, but also differences in their levels of expended effort. In short, scores on a test obtained in a low-stakes setting may be indicative of not only the students’ abilities but their level of motivation as well. Inferences about students’ abilities may then be questionable at best and inaccurate at worst (Haladyna & Downing, 2004; Wise & Smith, 2011). Motivation tends not to be a concern for high-stakes tests (e.g., college admission tests, college course tests, licensure exams). Students tend to expend high effort when completing tests that have personal consequences related to test performance (e.g., being admitted to college, passing a course, getting licensed).

Low-stakes tests are often used for institutional accountability mandates (Liu, 2017). These test scores are used to inform decisions that affect the university (Wise & Smith, 2011). State funding and university accreditation is dependent on the university’s
ability to show that students have changed in a positive way as a result of completing college curriculum. Using test scores obtained in a low-stakes contexts can be problematic for this purpose, as researchers have empirically demonstrated that test scores and estimates of the value-added of a college education computed from them can be attenuated if tests are low stakes (Finney, Sundre, Swain, & Williams, 2016; Wise & Smith, 2011). Thus, although low-stakes tests have low consequences of performance for the students completing these tests, the consequences of poor performance can be quite high for universities.

Given the empirical research demonstrating that motivation is important to consider when interpreting test scores from low-stakes tests, the Standards for Educational and Psychological Testing (American Educational Research Association, American psychological Association, & National Council for Measurement in Education, 2014) recommend motivation be reported to make accurate inferences from the test scores. Specifically, Standard 13.9 states, “In evaluation or accountability settings, test results should be used in conjunction with information from other sources when the use of additional information contributes to the validity of the overall interpretation” (AERA, APA, & NCME, 2014, p. 213).

In addition to reporting students’ motivation to aid test score interpretations, it is useful to study what constructs influence the amount of effort students expend. Identifying antecedents of motivation is important because the antecedents may be malleable; thus, testing practitioners could manipulate them to increase students’ motivation. Using Expectancy-Value theory as a frame to understand motivation, several studies suggest that students’ perceived value of a test and expectancy to succeed on a
test influence the level of students’ motivation. Using the Demands-Capacity model as a frame to understand motivation, studies suggest that test characteristics (e.g., test difficulty, test length) and students’ internal factors (e.g., level of proficiency, competitiveness, test anxiety) influence the level of students’ motivation. Using Control-Value theory as a frame to understand motivation, studies suggest that students’ emotions (e.g., boredom, anger, joy) influence the level of students’ motivation. In the present study, I use these theories to design and test robust models that estimate the effects of perceived value, self-efficacy, and test-emotions on students’ motivation and test performance. Below I provide a brief review of each theory prior to specifying the longitudinal panel models tested in the current study.

**Expectancy-Value Theory Applied to Examinee Effort**

Expectancy-Value (EV) theory provides an explanation of motivation for a variety of tasks in achievement settings (Wigfield & Eccles, 2000). According to EV theory (see Figure 1), the primary determinants of student’s motivation are 1) the expectancy of how well the student can perform the achievement task, and 2) the student’s subjective value of the achievement task (Flake, Barron, Hulleman, McCoach, & Welch, 2015; Wigfield & Eccles, 2000). Furthermore, students’ expectancies and subjective task values are hypothesized to affect students’ motivation via multiplicative effect. However, researchers note that the multiplicative effect has been often ignored and empirically not tested (Nagengast et al., 2011). Thus, empirical evidence supporting the multiplicative effect of expectancy and task value on students’ motivation is scarce (Guo et al., 2016).
Moreover, in the context of low-stakes testing, researchers frequently disregarded students’ expectancies when modeling examinee effort (Cole et al., 2008). In accountability testing contexts (K-12 or higher educations), students may not receive feedback on how well (or poorly) they performed on the low-stakes tests. Thus, before such tests, students may not have accurate expectancies about success on these tests, as they have never received such feedback. Moreover, to have accurate expectancies about a test, students need to know what the test will entail (e.g., content, difficulty, length, item type), which may not be the case in low-stakes testing contexts. Thus, if the goal is to model the impact of expectancies on examinee effort, time is needed before the test begins to expose students to examples from the test, which would allow them to formulate expectancies about their ability to succeed on the test. This time may not be budgeted in operational testing sessions.

Only recently, researchers have measured and modeled students’ expectancies to succeed on low-stakes tests when predicting examinee effort and test performance (Barry & Finney, 2016; Penk & Richter, 2017; Penk & Schipolowski, 2015). In one study, students’ expectancies did not relate to the initial effort level (i.e., effort measured after the first of four tests) and the rate of change in examinee effort throughout a low-stakes testing session (Barry & Finney, 2016). However, expectancies were measured after the testing session was over, which, although this allows students exposure to the test, it is problematic in that the temporal ordering of expectancies’ effect on examinee effort is violated. That is, expectancies should be measured before examinee effort. When expectancies were measured before a test, the initial level of expectancies (i.e., expectancy measured before the test) related to initial level of examinee effort (i.e., effort
measured before the test), after controlling for test importance (Penk & Richter, 2017) and after controlling for test importance, interest, and anxiety (Penk & Schipolowski, 2015). Furthermore, the average rate of change in students’ level of expectancies related positively to the average change in examinee effort (i.e., as students decreased in expectancies at a greater rate throughout the test, they also decreased in expended effort at a greater rate). Lastly, students’ expectancies measured before a test related positively to the performance on the test, after controlling for test importance, interest, anxiety, and intended effort measured before a test (Penk & Schipolowski, 2015) and after controlling for initial test importance (i.e., importance measured before the test), initial effort (i.e., effort measured before the test), domain self-concept, and several demographic variables (i.e., gender, school track, socioeconomic status, immigration background, and ability; Penk & Richter, 2017). Although expectancies were measured before the test in the two Penk and colleagues’ studies, students were not exposed to example test items. Thus, students may not have had necessary information about the test to inform accurate ratings of expectancies.

According to EV theory, students’ subjective value consists of four components: interest, usefulness, importance, and cost. Interest refers to the intrinsic value that students experience as they complete an achievement task. Usefulness refers to the extrinsic value or utility of the achievement task. Importance, or attainment value, refers to personal relevance of the achievement task. Lastly, cost refers to the time, effort, or emotional demands of the task (e.g., anxiety). Theoretically, interest, usefulness, and importance should have a positive effect on students’ motivation whereas cost should have a negative effect on students’ motivation.
Empirical evidence supports the hypothesized positive relationship between students’ perceived value of a test and examinee effort (e.g., Barry & Finney, 2016; Cole et al., 2008; Finney, Myers, & Mathers, 2018; Penk & Richter, 2017; Penk & Schipolowski, 2015; Thelk, Sundre, Horst, & Finney, 2009; Zilberberg, Finney, Marsh, & Anderson, 2014). Given this relationship, researchers have attempted to manipulate students’ perceived value of a test to increase students’ effort, which in turn, should lead to a higher test performance (Finney et al., 2016; Hawthorne, Bol, & Pribesh, 2015; Liu, Rios, & Borden, 2015; O’Neil, Abedi, Miyoshi, & Mastergeorge, 2005). The results from these studies are mixed: some manipulations increased students’ expended effort on tests, others failed.

Many of the studies referenced above that investigated examinee effort using EV theory have methodological issues. First, studies conducted using cross-sectional data (e.g., Cole et al., 2008; Finney et al., 2018; Liu et al., 2015) present concerns about temporal ordering of constructs involved in the causal EV model (see Figure 1). According to EV theory, perceived value of a test and expectancy to succeed should affect subsequent motivation, which then affects subsequent test performance. In many studies, perceived value, expectancy, and motivation are measured at the same time, thus violating the proper ordering of the causal mechanism. Tests of causal theories, such as EV theory, necessitate a longitudinal research design with appropriate time lags between measures of constructs (Preacher, 2015). The lags are necessary for antecedents to have time to impact subsequent behaviors and attitudes. If the constructs are measured at the same time (i.e., concurrently) then the causal hypothesis would be that “antecedents” affect behaviors and attitudes instantaneously.
Second, measures of these constructs are often gathered upon test completion (e.g., Barry & Finney, 2016; Cole et al., 2008). In other words, test-taking importance, expectancy, and motivation are assessed after performance on a test is already determined. Assessing value, expectancy, and motivation after the test can result in self-protective bias (Myers, 2017). That is, if students perceive that they performed poorly on the test, they may choose to report low levels of expectancy, value, and effort, in order to preserve self-esteem.

Third, although studies exist where students’ test value, expectancies, and motivation were measured longitudinally (e.g., Penk & Richter, 2017; Penk & Schipolowski, 2015), no auto-regressive effects were estimated in these models. Auto-regressive effects are the effects of a construct on the same construct measured at a subsequent time (e.g., effect of motivation before a test on motivation after the test). Longitudinal models without auto-regressive effects are misspecified to some degree, because the best predictor of a construct is typically an earlier measurement of that same construct. Auto-regressive effects serve as a statistical control with respect to the effects of other predictors in the model. That is, when the effects of other variables are estimated, auto-regressive effects are taken into account, thus, strengthening the argument for causality with respect to these other variables (Cole & Maxwell, 2003; Jose, 2016; Preacher, 2015).

Fourth, as noted above, a prominent issue with many of these studies (with the exception of Barry & Finney, 2016; Penk & Richter, 2017; Penk & Schipolowski, 2015) is the exclusion of expectancy and the expectancy-value interaction. By omitting student
expectancies or the interaction between expectancy and value from models predicting examinee effort, the effects of perceived value of a test might be biased.

Given the lack of longitudinal studies in which constructs are specified in a proper temporal order with auto-regressive effects included, coupled with the lack of modeling the expectancy component (and the interaction between expectancy and value) of EV theory, the empirical support of EV theory in low-stakes testing may be premature. In the current study, I will address these methodological limitations.

**Demands-Capacity Model of Examinee Effort**

Wise and Smith put forth a new model of examinee motivation that built off of EV theory (Smith & Holterman ten Hove, 2010; Wise & Smith, 2011, 2016). According to the Demands-Capacity model, two primary constructs determine the amount of effort examinees will be expend on a test: the resource demands of the test and the effort capacity of the examinee (Wise & Smith, 2011, 2016). The resource demands of a test, and more specifically of each item, are determined by test characteristics such as item difficulty, the length of each item, or the number of items on a test. Resource demands represent a fixed quantity (for a given item) that an examinee needs to expend to correctly answer an item. Resource demands can vary from one item to another item, which is in alignment with EV theory, in which antecedents of student motivation (i.e., students’ expectancy and value) may also vary from one test to another.

The effort capacity of each examinee refers to how much effort students are willing to expend on the test. According to the Demands-Capacity model, variables that affect the level of effort capacity include the test’s consequences for the examinee, time pressure to complete the test, examinees’ level of preparation for the test, examinees’
expectations about the demands of a test, and examinees’ emotions during the test. Effort capacity is theorized to change throughout the test. For example, as the test progresses, students may get fatigued, thus reducing their effort capacity for completing subsequent items. Effort capacity is related to EV theory in that effort capacity is a function of examinee’s expectancy and value. That is, examinee’s expectation about the test and level of proficiency will influence examinee’s expectancies to succeed. Additionally, test consequences will influence the perceived value of the test.

Effort capacity is also theorized to be influenced by test anxiety. Likewise, researchers using EV theory to understand expended effort have conceptualized anxiety as an emotional cost to complete a test (e.g., Penk & Schipolowski, 2015). Anxiety, present in both EV theory and the DM model, is the most studied emotion in educational literature. However, the Control-Value theory of Achievement Emotions asserts that students experience many other emotions (e.g., pride, anger, joy) in achievement situations. In this study, five test emotions (i.e., anger, joy, boredom, worry, and pride) were measured and their effects on examinee effort were modeled, as specified below.

**Control-Value Theory: Explicit Incorporation of Achievement Emotions**

Control-Value theory (CV) of Achievement Emotions describes the development of achievement-related emotions and how these emotions affect achievement outcomes via cognitive-motivational constructs (Pekrun, 2006, 2007, 2017; see Figure 2). According to CV theory, two situational appraisals of an achievement situation will largely determine what emotions will be experienced: perceived control and subjective value. Perceived control refers to the degree that students’ expect to perform and succeed on an achievement task. Perceived control can be differentiated into three types of
expectancies: action-control expectancies, action-outcome expectancies, and situation-outcome expectancies (Pekrun, 2006). Subjective value refers to the importance of a task to students. Subjective values can be differentiated into three types of value: intrinsic, achievement, and extrinsic. Notice that the two situation appraisals (i.e., perceived control and subjective value) that determine subsequent achievement emotions are similar to the determinants of motivation in EV theory (expectancy to succeed and subjective value). Both theories are similar in that they explain the development of motivation; however, subtle differences exist between the two theories (see Putwain et al., 2018). For example, in EV theory, the situational appraisals of perceived value and expectancy are hypothesized to affect students’ motivation directly (Figure 1), whereas these effects on motivation in CV theory are indirect via emotions (Figure 2).

According to CV theory, several constructs cause situational appraisals (Figure 2). The purpose of this study is not to investigate how constructs affect situational appraisals; however, they are worth mentioning. Task demands, autonomy support in achievement settings, students’ goal structures, feedback, and consequences for performing achievement tasks are several examples that are hypothesized to influence situational appraisals (Pekrun, 2007). Notice that some of constructs, which fall into the “Environment” category in Figure 2, are similar to variables that are hypothesized to influence effort capacity and resource demands in the Demands-Capacity model. For example, in the Demands-Capacity model, test (and individual item) demands, such as item difficulty, influence resource demands, whereas test consequences influence how much effort students will put forth in answering subsequent test items (Wise & Smith, 2011, 2016). In CV theory, these constructs are theorized to affect situational appraisals
(i.e., expectancies and values), which in turn affect emotions, which then affect students’ effort and other cognitive-motivational variables. Thus, CV theory specifies how the effects of these “environmental” variables unfold, whereas the Demands-Capacity model simply acknowledges the importance of these variables.

In CV theory, environmental variables are hypothesized to impact the situational appraisals, which then impact achievement emotions. These achievement emotions are categorized along three dimensions: object focus, valence, and activation (see Table 1). Object focus refers to the emotions that develop with respect to an activity or an outcome. For example, test anxiety is an activity emotion because it develops in response to the activity of testing. Outcome emotions can be either prospective or retrospective. For example, students may experience hope (i.e., prospective outcome emotion) before an outcome is known to them, or they may experience shame or pride (i.e., retrospective outcome emotions) after an outcome has occurred. The second dimension is valence, which refers to whether the emotions are pleasant or unpleasant. Positive emotions (i.e., enjoyment) are defined as being pleasant to students, whereas negative emotions (i.e., anger) are defined as being unpleasant to students. The third dimension is activation. Activation refers to the physiological arousal that emotions produce. Emotions can be either activating or deactivating. For example, students who experience activating emotions (i.e., anxiety) may experience increased heart rate. In contrast, students who experience deactivating emotions (i.e., relief) will not experience elevated heart rate.

CV theory explains how these achievement emotions affect task performance via several cognitive-motivational variables: cognitive resources, learning strategies, self-regulation, and motivation. The distinction between types of emotions is useful when
examining the effects of emotions on these cognitive-motivational variables (see Table 2). Positive emotions (e.g., enjoyment or hope) relate positively to motivation, flexible learning strategies, and self-regulation (Pekrun et al., 2002; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). Negative emotions (e.g., anger or anxiety) relate negatively to motivation, flexible learning strategies, and self-regulation (Pekrun et al., 2002; Pekrun et al., 2011).

In addition to unidirectional effects between academic environment, situational appraisals, achievement emotions, cognitive-motivational variables, and task performance, CV theory posits that all these constructs are reciprocally related (Pekrun, 2007). Several studies have investigated the reciprocal relationship between emotions and task performance. Positive emotions tend to positively affect task performance (commonly measured by test scores), whereas negative emotions negatively affect task performance; in turn, task performance positively affects subsequent positive emotions, and negatively affects subsequent negative emotions. (Gibbons, Xu, Vilafane, & Raker 2018; Pekrun, Hall, Goetz, & Perry, 2014; Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017; Putwain, Becker, Symes, & Pekrun, 2018; Reeve, Bonaccio, & Winford, 2014).

The focus of this study is on the application of CV theory in a testing context to better understand the antecedents of examinee effort. Studies have suggested that different test emotions are experienced before, during, and after a test (Daniels & Gierl, 2017; Goetz et al., 2007; Pekrun et al., 2004; Putwain et al., 2018). The theorized reciprocal effects have also been examined in a testing context. Positive (joy, hope, and pride) and negative (anger, shame, and anxiety) test emotions had effects on test
performance and, in turn, test performance had a positive effect on subsequent positive emotions (pride, relief) and a negative effect on subsequent negative emotions (anger, shame) (Reeve et al., 2014). No studies have examined the effects of test emotions on subsequent examinee effort or the reciprocal effects of examinee effort on subsequent test emotions. Thus, the present study will address this gap in the literature.

**Purpose of the Current Study**

The purpose of the current study is two-fold. The first purpose is to employ a rigorous and appropriate design to test the EV hypothesis that expectancy and value impact subsequent examinee effort. The majority of studies examining examinee effort in a low-stakes testing context have utilized a cross-sectional research design (e.g., Cole et al., 2008; Finney et al., 2018; Liu et al., 2015). If the goal is to estimate mediation effects specified by EV, cross-sectional data violates the temporal ordering of the variables. Limited inferences can be drawn from cross-sectional studies that find that perceived value and expectancy relate positively to expended effort and expended effort relates positively to performance on the test (Preacher, 2015). Relations deemed indirect in nature may instead be spurious. In the current study, I gathered stronger evidence to evaluate mediation by employing a longitudinal design that incorporated both time lags between constructs involved in mediation and auto-regressive effects.

Moreover, the majority of studies framed using EV theory did not include the main effect of expectancy on effort (with the exception of Barry & Finney, 2016; Penk & Richter, 2015, Penk & Schipolowski, 2017) or the interaction between expectancy and value. By excluding these effects, the models do not fully represent or test EV theory. In the current study, self-efficacy ratings will be collected. Given adequate measurement,
the direct and indirect effects of expectancy on subsequent examinee effort will be estimated.

The second purpose of the current study is to examine the role of test emotions in the development of examinee effort. I specified and tested competing models that represent EV and CV theory (detailed model specifications and comparisons are articulated in Chapter 3 under the Planned Analysis section). The CV Theory Model A (Figure 12) specifies the effects of test emotions on subsequent examinee effort and the effects of examinee effort on subsequent test performance. As explained above, CV theory encompass EV theory, thus some components of EV theory are present in the CV model. Namely, examinee effort is theorized to affect subsequent test performance. However, compared to the EV Theory model, the antecedents of examinee effort are different in the CV Theory model. According to CV theory (Figure 2), test emotions directly affect subsequent examinee effort. Perceived value of a test and expectancy affect effort indirectly via test emotions. Thus, the CV Theory Model (Figure 12) will be tested and compared to EV Theory Model (Figure 8).

In addition to comparing EV and CV Theory Models, a set of CV Theory models (Figures 13-16) will be estimated to test reciprocal effects between emotions and motivation. According to CV theory, test emotions not only affect subsequent examinee effort, but examinee effort also affects subsequent test emotions. Given the longitudinal design, the reciprocal effects of test emotions and examinee effort can be tested. Moreover, CV theory specifies reciprocal effects between all constructs (Figure 2). Thus, the reciprocal effects between perceived value, test emotions, effort, and performance
will be tested. The specific models and associated research questions are presented in Chapter 3 after providing additional detail about the modeling technique.
Chapter 2: Literature Review

The literature review will summarize research on two broad concepts that are central to the current study: achievement emotions and mediation. In the first section, I review the literature on Control-Value (CV) theory of achievement emotions. First, I define achievement emotions and describe how they are different from other emotions. Second, I categorize achievement emotions according to CV theory. Third, I explain what constructs are hypothesized to influence achievement emotions. Fourth, I describe how emotions influence achievement indirectly via several cognitive-motivational constructs. Then, I describe how CV theory hypothesizes reciprocal relationships between achievement emotions and all the constructs involved in the theory (see Figure 2).

In the second section, I define mediation and its utility when articulating hypotheses regarding the relation between variables. Second, I describe three common approaches or designs used to gather data for mediation analysis. I note advantages and disadvantages of each design. Third, I define three concepts of change and describe how each research design can examine change. Fourth, I provide rationale for longitudinal designs when modeling mediation.

Achievement Emotions

Achievement emotions are defined as emotions that pertain to achievement experiences such as studying for an exam, participating in class, or taking a test (Pekrun, 2006, 2007). The difference between achievement emotions and other types of emotions, such as social emotions and topic emotions, is the object focus. Social emotions are described as emotions that arise because of social situations. That is, the focus of emotions is other people or the social contingencies for a specific situation. For example,
one student might be jealous of another student’s grade on a homework. The object of jealousy in this situation is another student and the other student’s performance. Topic emotions arise because of certain feelings towards a specific topic or content of a lecture. For example, racial or ethnic oppression may be a sensitive topic to some groups, which may affect how students are feeling during a class in which the sensitive topic is introduced. In contrast to achievement emotions, topic emotions do not relate to learning directly; however, they may be related to engagement (Pekrun & Perry, 2014). The object focus of achievement emotions are achievement activities or achievement outcomes. For example, anxiety felt during a test is an achievement emotion because emotion is stemming from fear of failing the test (i.e., achievement outcome). Social and topic emotions (among other types of emotions) are present in achievement settings but CV theory of achievement emotions (theory used as the basis for the current study) does not encompass them.

The studies of achievement emotions have increased dramatically from the 1950s to now. Most of the research then and now focuses on anxiety, and, more specifically, test anxiety. The rich literature on test-anxiety suggests that test-anxiety tends to have detrimental effects on achievement (Zeidner, 2014). However, anxiety is not the only achievement emotion that is experienced by students. Pekrun, Goetz, Titz, and Perry (2002) conducted multiple studies in which they found that other emotions are experienced at high frequency in achievement situations. Specifically, they examined what emotions university students felt when completing homework, preparing for a class, or taking a test. Anxiety was experienced most frequently during activities; however, anger, pride, enjoyment, and boredom were reported quite often as well, suggesting the
need for further exploration of these “other” emotions and their effects on achievement. Since the initial studies in the early 2000s, researchers have completed multiple studies building and testing the theory of achievement emotions.

**Control-value theory of achievement emotions.** CV theory (see Figure 2) hypothesizes relationships between achievement emotions, the antecedents of emotions, and outcomes of achievement emotions. Most of the research testing the predictions of CV theory has been spearheaded by Reinhard Pekrun and his colleagues. In fact, Pekrun and his colleagues developed CV theory based on their previous research on cognitive and motivational effects on achievement emotions. Thus, this review is largely based on work produced by Pekrun’s research teams (e.g., Pekrun, 2006, 2017; Pekrun & Perry, 2014; Pekrun et al., 2002).

According to CV theory, achievement emotions can be categorized using three dimensions: object focus, activation, and valence (see Table 1). The object focus can be the outcome prospectively, the outcome retrospectively, or the activity. For the outcome emotions, students experience emotions related to the outcome of an achievement activity. For example, students may experience anxiety when preparing for an important exam (i.e., prospective outcome-related emotion). Additionally, students may experience relief or shame after an exam (i.e., retrospective outcome-related emotions). The outcome in both situations is the exam performance. Emotions can also occur with respect to an achievement activity. For example, students may be bored when completing easy homework or students may be angry during a test.

Achievement emotions can also be differentiated by the type of activation. Activation refers to the physiological arousal an individual experiences in regards to an
achievement outcome or activity. Achievement emotions can be activating or
deactivating. Activating emotions, such as anger or anxiety, produce physiological
activation. A common example of a physiological symptom produced by high test
anxiety is increased cortisol levels (Pekrun et al., 2002), which is a common way to
measure physiological arousal. Other indications of high physiological arousal are high
heart rate, high blood pressure, and high respiration rate. Deactivating emotions, such as
relief or boredom, are defined as emotions that produce low physiological activation.
Deactivating emotions are the opposite of activating emotions, producing the opposite
pattern of symptoms (e.g., low heart rate, low blood pressure, low respiration rate).

Achievement emotions can also be differentiated by valence: positive versus
negative emotions. Positive emotions, such as enjoyment and pride, are pleasant to the
individual. Negative emotions, such as anger and boredom, are unpleasant to the
individual.

When crossing the three dimensions (see Table 1), each emotion can be
specifically categorized (Pekrun & Perry, 2014). It is hypothesized that emotions of the
same category have similar antecedents and similar outcomes (see Table 2 and Figure 2).
Researchers have shown that positive activating test emotions (e.g., joy, pride) relate
positively to the antecedents of achievement emotions, which are perceived control and
value, whereas negative activating test emotions (e.g., anger, anxiety) relate negatively to
perceived control and value (Pekrun, Goetz, Perry, Kramer, Hochstadt, & Molfenter,
2004). Researchers have also supported the negative relationship between negative
activating emotions (e.g., anger, anxiety) and the outcome of motivation to learn and
supported a positive relationship between negative activating emotions and the outcome
of irrelevant thinking (Pekrun et al., 2002). Additionally, positive activating test emotions (e.g., joy, pride, hope) relate positively to achievement outcomes (e.g., performance on an exam), whereas negative deactivating test emotions (e.g., hopelessness) relate negatively to achievement outcomes (Pekrun et al., 2004). In the next section, I review CV theory and the antecedents of achievement emotions (i.e., appraisals of control and appraisals of value) in greater detail given this research informs the specification of the models tested in the current study.

**Antecedents of emotions.** According to CV theory, situational appraisals are hypothesized to produce different emotions (Pekrun, 2006). Building on attributional theory (Weiner, 1985), CV theory posits that two situational appraisals cause achievement emotions: the appraisal of perceived control over a situation, and the appraisal of value of a situation (see Figure 2).

**Appraisals of control.** Perceived control is defined as the appraisals of control over actions and outcomes which determine the subjective expectation or confidence to achieve success on achievement task (Pekrun, 2006; Pekrun & Perry, 2014). According to CV theory, perceived control consists of three types of causal expectancies: action-control expectancies, action-outcome expectancies, and situation-outcome expectancies (Pekrun, 2007). Action-control expectancies are the beliefs about how well an action can be performed. For example, a student might have an expectation that s/he is able to take detailed notes in class. The action-control expectancy is that the student is simply able to take notes and it does not refer to what outcome the action of taking notes might lead to. Self-efficacy is a term that is often used to describe action-control expectancies (Bandura, 1977). Action-outcome expectancies are beliefs that a particular action will lead to a
desirable outcome. Compared to the action-control expectancies, the focus of action-outcome expectancies are on the outcome. For example, a student may believe that taking detailed notes will lead to remembering the lecture material. That is, the student may hold a belief that s/he is able to take the detailed notes, but also expects the notes will lead to a successful outcome. In other words, the student’s action will result in an outcome. The situation-outcome expectancies are beliefs that one does not have control over the outcome. Instead, the situation the student is in will determine the outcome. For example, a student that holds situation-outcome expectancies may believe that s/he will obtain a good grade due to something external to the student (e.g., a lenient teacher). Pekrun (2007) argues that situation-outcome expectancies are the least relevant of the three expectancies, because, in achievement settings, students often have a level of control.

The combination of all three situational control appraisals will dictate what emotions will be experienced before engaging in a task. Note, perceived control to perform well on an exam is separate from the actual performance on the exam. That is, the appraisal of perceived control, rather than the actual level of success, will dictate what emotions will be experienced before and during the achievement activity. Once students receive feedback about their performance on a task, appraisals of control will be updated, which may lead to different appraisals for a given task in the future.

To illustrate how perceived control determines achievement emotions, consider a student who believes that s/he has full control in preparing for a test (action-control expectancy). That is, the student knows how to study for a test, knows what material will be emphasized, and based on previous experience taking tests, knows that s/he is capable of completing a test. This student may experience joy and hope when studying for the
test. In addition, if the student believes that the preparation will lead to a desirable grade (action-outcome expectancy), then this student may feel hopeful about the performance on the test and experience low levels of anxiety. In contrast, consider a student who feels that in a particular class, s/he cannot attain a good grade because the instructor uses an ambiguous rubric to score the exam (situation-outcome expectancy). This student will likely feel hopeless or anxious because attaining the success (or failure) on the test is out of student’s control.

**Appraisals of value.** In addition to perceived control, the value one places on achievement success or failure is important (Pekrun, 2006). According to CV theory, there are two types of values in achievement setting: intrinsic and extrinsic. Intrinsic value refers to the importance of performing an activity for the sake of engaging in an activity, without placing much attention on the outcome. For example, students might perceive attending class as important (i.e., having high intrinsic value) because the material is engaging and interesting. Extrinsic value refers to the importance of an activity because attaining a favorable outcome is desirable. For example, students may perceive class as important because it will lead to getting a rewarding job. Thus, value may be in regard to an activity itself (e.g., attending class) or in regards to an outcome of the activity (e.g., doing well in class may lead to getting a job offer). Recent developments of the theory have proposed and empirically tested that, in addition to intrinsic and extrinsic (or utility) values in achievement settings, students may have achievement value (Putwain et al., 2018). Achievement value is defined as the importance of doing well on a task (e.g., attaining good or poor grade) for either intrinsic or extrinsic reasons. Thus, achievement value has elements of both intrinsic and extrinsic
values. That is, students may have high achievement value because they find positive outcomes rewarding but they also enjoy engaging in the activity itself. Importantly, when researchers tested interactions between perceived control and a type of value (intrinsic, extrinsic, or achievement) on subsequent boredom and enjoyment, they found different effects of the interaction, suggesting that the three types of values are distinct and have different effects on subsequent emotions (Putwain et al., 2018). Specifically, they found a significant positive interaction effect of perceived control and achievement value on subsequent enjoyment. Additionally, intrinsic and utility value had positive effects on subsequent enjoyment. The interaction between perceived control and intrinsic value had a negative interaction effect on subsequent boredom.

Whereas including and modeling achievement value in CV theory is a new development, researchers in the area of educational psychology have hypothesized separate types of values for decades (Eccles et al., 1983). CV theory of achievement emotions was developed by using control-value appraisals to explain achievement emotions. However, the situational appraisals were borrowed from the Expectancy-Value (EV) theory of motivation (Wigfield & Eccles, 2000). That is, the control-value appraisals in CV theory aligns with expectancy-value appraisals in EV theory, even though the purpose of the appraisals in each theory is different. The appraisals in CV theory explain why achievement emotions develop, whereas the appraisals in EV theory explain why motivation develops (Putwain et al., 2018).

According to EV theory, four types of subjective values exist: attainment value, intrinsic value, utility value, and cost. There are similarities between the conceptualization of values in EV theory and CV theory (Table 3). Both theories
recognize and define intrinsic value in similar ways. The utility value in EV theory aligns with extrinsic value in CV theory. Theoretically, both of these values focus on the outcome – the achievement outcome in CV theory, and the usefulness of task (in contrast to the task itself) in EV theory. That is, in EV theory, utility value refers to how useful engaging in an activity may be for attaining individual’s goals. In CV theory, extrinsic value refers to the importance of an achievement task to achieve an desirable outcome. Thus, theoretically, both, the utility value in EV theory and extrinsic value in CV theory are referring to the same construct (Putwain et al., 2018). While intrinsic and extrinsic/utility values are similar in both CV and EV theory, the achievement value in CV theory is different from attainment value in EV theory. Achievement value in CV theory refers to the importance of doing well on an achievement task for either intrinsic or extrinsic reasons. Attainment value in EV theory refers to importance of achieving an desirable outcome for preserving one’s positive self-identity (Putwain et al., 2018). The last component in EV theory for value is the cost of performing a task. Cost refers to what individuals have to give up in order to engage in an activity and the anticipated effort required to perform an activity. Historically, researchers have ignored the role of cost in EV theory (Cole et al., 2008; Wigfield & Eccless, 2000). Only recently, researchers included cost in their conceptualization of EV theory (Flake, Barron, Hulleman, McCoach, & Welch, 2015). Cost is not included in CV theory.

When discussing the antecedents of emotions and specifically appraisals of values, it is useful to review the studies that have examined the impact of values in low-stakes testing context. That is, EV theory has been applied to low-stakes test contexts to predict test-taking motivation and performance on a test and appraisals of value play a
prominent role in this research (e.g., Cole, Bergin, & Whittaker, 2008; Penk & Schipolowski, 2015, Penk & Richter, 2017). In one study using cross-sectional data design, researchers found that three different types of subjective values (usefulness/utility, interest/intrinsic value, and importance/attainment value) predicted test-taking motivation (operationalized as test-taking effort) differentially, which then had a positive effect on performance on four different subject (i.e., English, Math, Science, Social Studies) tests (Cole et al., 2008). Researchers found positive effects of each type of value on test-taking effort in all but one instance. Interest had a negative effect on test-taking effort for English test. The standardized effects of value on effort for each subject test ranged from .41 to .53, indicating moderate effects. The standardized effects of test-taking effort on performance for each subject test ranged from .41 to .53, indicating moderate effects. This study was limited in that researchers did not measure or model expectancy to perform well on the test (as EV theory suggests); however, researchers showed that different values have different effects on test performance via test-taking effort.

Using three-wave longitudinal research design, Penk and Richter (2017) measured test-taking importance (attainment value), probability of success (self-efficacy or expectancy to succeed), test-taking effort, and performance on mathematics test. Test-taking importance, probability of success, test-taking effort were measured before the test, after completing half of the test, and at the end of the test. They found that initial levels of probability of success and test-taking importance had an effect on initial level of intended test-taking effort. Furthermore, how students changed throughout the test in probability of success and test-taking importance had a positive effect on how students
changed in test-taking effort. The initial level of test-taking effort had a positive effect on test performance, whereas the effect of how students changed on effort throughout the test on test performance was not statistically significant. Interestingly, the initial probability of success and how students changed on it had a direct effect on test performance, in addition to the indirect effect via initial level of test-taking effort. The direct effects of initial test-taking importance and how students changed on test-taking importance on test performance were not statistically significant. Although this study by Penk and Richter included probability of success, which addresses the expectancy component of EV theory, they only modeled one type of subjective value (i.e., importance).

A study more aligned with EV theory than the two previously mentioned studies was conducted by Penk and Schipolowski (2015). They measured probability of success (expectancy to succeed), three types of subjective values (importance (attainment value), interest (intrinsic value), and anxiety (cost)), test-taking effort (motivation), and performance on a mathematics test. Noteworthy for this study, the researchers operationalized anxiety as perceived cost of completing a test. All variables except performance on the test were measured before the test and after the test. This study replicated the moderate effect (Cole et al., 2008) of test-taking effort on performance (standardized effects were .33 and .30 for effort measured before and after the test, respectively). Results indicated that probability of success, importance, interest, measured both before and after the test had positive effects on test-taking effort. Anxiety measured before the test had a negative effect on subsequent effort. Anxiety measured after the test did not have an effect on subsequent effort measured after the test.
Several conclusions can be drawn from the studies summarized above. First, the three studies conducted using low-stakes tests evidenced that test-taking motivation had a positive effect on performance on a test. Second, taken together, the studies show all components (expectancy to succeed and different type of values) of EV theory are important to consider when predicting test-taking motivation. Third, longitudinal studies remain a rarity when testing EV theory. That is, expectancies and values are still measured concurrently with motivation, which violates the time precedence of ordering of these variables. Fourth, according to EV theory, the situational appraisals of expectancy to succeed and subjective values should have a multiplicative effect on subsequent motivation. In other words, an interaction between the two components of EV theory should be modeled. To my knowledge, only one study has examined the multiplicative effect of expectancy and value on motivation (Guo et al., 2016). In this study with 9th graders, researchers measured intrinsic value, attainment value, utility value, and cost. Expectancies were defined as students’ self-concept in mathematics. The results uncovered a significant effect of interaction between global value (higher-order latent construct, in which four types of values were combined) and students’ self-concept. However, the study utilized cross-sectional data; thus, interpretation of the significant interaction effect is questionable.

The multiplicative effect of situational appraisals on emotions. Recall that CV theory specifies that perceived control and value will dictate if and what achievement emotions will develop (see Figure 2), unlike EV theory which specifies these two appraisals impact motivation directly (Pekrun, 2006, 2007). Importantly in CV theory, the combination refers to the interaction (or multiplicative effect) between level of
perceived control and subjective value when predicting achievement emotions (Putwain et al., 2018). Furthermore, according to CV theory, subjective value is framed with respect to success or failure for prospective and retrospective outcome emotions. For example, if a student asserts that an outcome is important (high intrinsic, extrinsic, or achievement value), experiencing success or failure on that outcome will dictate what emotions will be felt. These appraisals of perceived control and value apply to all three different foci of emotions: prospective outcome, retrospective outcome, and the activity emotions.

Table 2 summarizes what emotions should arise in each situation depending on the levels of control and values (Pekrun, 2006). It is important to note that CV theory does not imply that appraisals of perceived control and values are the only factors that spark achievement emotions. There are certainly other situational appraisals, such as achievement goals that affect emotions, however perceived control and subjective values are considered to be the most important.

First, consider prospective outcome emotions (e.g., hope, anxiety). If a student perceives that s/he has high control (i.e., high self-efficacy) in attaining a high grade on an important (i.e., high value) test, then either anticipatory joy or anticipatory relief may be experienced before a test. When value is high, which emotion will occur will be dependent on whether the student focuses on success (anticipatory joy) or failure (anticipatory relief). If the student feels moderate level of control, hope for success (attaining high grade) or anxiety of failing (attaining low grade) will be experienced before a test. Student who believes s/he have no amount of control will experience hopelessness, regardless if the focus is on success or failure. If the outcome of the test
(high or low grade) is not regarded as important (i.e., low intrinsic, extrinsic, or achievement value), then a student may experience relief if the perceived control is high or experience no emotions before the test if perceived control is low.

Now, consider retrospective outcome emotions (e.g., joy, sadness). For retrospective outcome emotions, perceived control is said to have a minimal effect on development of these emotions because the outcome has already happened and the current level of perceived control cannot affect the outcome, only the prior level of perceived control can affect the outcome (Pekrun, 2006). Which retrospective outcome emotions occur will be determined by causal attribution of who exerted the control: the individual, others, or the causal agent is irrelevant. For example, consider a student who received a grade on a test and values this outcome (i.e., high value). If the student perceives that s/he is responsible for that grade, pride will develop if the test grade is high (success) or shame will develop if the test grade is low (failure). If a student perceives that other people are responsible for him/her receiving a high grade on the test, gratitude will develop. If a student perceives that other people are responsible for him/her receiving a low grade, anger will develop. If it does not matter who caused high or low grade, a student will experience either joy or sadness following a test. Notice in Table 2, there is no low value for retrospective outcome emotions. If value is low then the intensity of pride, gratitude, shame, or anger will be reduced (Pekrun, 2006).

Now, consider activity or task emotions (e.g., enjoyment, anger). In contrast to outcome emotions (prospective or retrospective), the activity itself will be evaluated. If an activity is important (high value) to the student and control is high, the student will experience enjoyment when engaging in the activity. For example, consider a student
who is studying for a class. If student feels in control of studying (high self-efficacy) and studying is an important activity (high value), the student will experience enjoyment during studying. If studying is not a valued experience because the student would rather play video games than study organic chemistry, the student will feel anger even if perceived control is high. However, if perceived control is low, thus the student feels s/he is not capable of studying, then student will experience anger if value is high or boredom if value is low.

**Antecedents to control and value appraisals.** Given perceived control and value impact subsequent emotions, it is helpful to briefly review what constructs impact appraisals of control and value (see Figure 2). Moreover, in the testing context, which is the focus of this thesis, much research has focused on how to influence perceived value in order to influence subsequent effort (Finney et al., 2016; Hawthorne, Bol, & Pribesh, 2015; Liu, Rios, & Borden, 2015; O’Neil, Abedi, Miyoshi, & Mastergeorge, 2005).

CV theory specifies the antecedents to appraisals of perceived control and value include environmental factors such as motivational quality, autonomy support, social expectations, and feedback and consequences of achievement (Pekrun, 2007, 2017). **Motivational quality** is a characteristic of the achievement environment. It refers to the prior history of achievement settings in which different motivation messages were delivered about the achievement. For example, in the classroom setting, the teacher may foster competition among the students and provide a reward for doing well on a test (Pekrun & Perry, 2014).

**Cognitive quality.** Cognitive quality is a characteristic of the achievement environment. It refers to the demands placed on students’ cognitive resources. For
example, the test’s characteristics may be more or less taxing on students’ cognition. Thus, students may prefer multiple-choice items rather than short essays (Smith & Holterman ten Hove, 2010), which is reflected by students expending less effort on essay items (Sundre, 1999). Cognitive quality includes several factors that are noted in the Demands-Capacity model of examinee effort. For example, test characteristics such as item length, item format (constructed response vs. multiple choice items), amount of reading required are all examples of variables that are theorized to impact students’ effort capacity in the Demands-Capacity model.

**Autonomy support.** Autonomy support is a characteristic of the achievement environment and refers to environments that facilitate or inhibit student’s perceived control. That is, some environments might help facilitate student’s sense of control by providing challenges and circumstances, in which students are taught to always exercise their agency. For example, instructors may structure homework assignments in a way that does not provide many instructions. Thus, to succeed, students have to seek out additional support and information, beyond what the homework assignment involve. In contrast, other environments may be structured in a way that asks students to carefully follow the instructions and provide all required materials. Thus, the only way to succeed is to follow the instructor’s plan. In one situation students may learn to actively and independently learn from additional material, while in the second situation, no additional efforts need to be invested.

**Social expectation.** Social expectation is a characteristic of the individuals in the achievement situation. It refers to expectations about achievement setting (i.e., behaviors, norms) that might be endorsed by parents, teachers, or peers, which then influence the
individual. For example, a student that comes from a family of mathematicians may have parental expectations that mathematics should be enjoyable and student should be naturally talented to do well in the course, thus the perceived control over a subject should be high.

**Feedback and consequences of prior achievement.** Feedback and consequences of prior achievement are characteristics of the individuals in an achievement environment. A student may have an extremely passionate teacher, who cares about their students’ learning. As a result, the teacher always tries to instill some confidence in students by providing positive feedback and highlighting the skills that they have. A student that receives such feedback over time might be taught to believe that exams are important to check the knowledge and skills that they have developed. Thus, the value component of achievement appraisals will be impacted. Conversely, negative feedback and consequences may lead to a student either devaluing a course or activity (maybe in order to preserve self-esteem) or to invest more effort in the future in order to avoid failure.

**Outcomes of emotions.** In addition to specifying antecedents of emotions, the three-dimensional model of emotions is useful when summarizing how emotions affect achievement (Pekrun, 2017). According to CV theory, the effects of achievement emotions on achievement are not direct (see Figure 2). Cognitive resources (e.g., types of information processing), interest and motivation (e.g., motivation to complete an achievement task), learning strategies (e.g., elaboration or rehearsal of material), and self-regulation (e.g., intrinsic or extrinsic) mediate the relationship between emotions and achievement. Thus, it is important to understand how those intermediate outcomes are
influenced by emotions (see Table 2). Positive activating emotions (e.g., enjoyment of a task) preserve cognitive resources, enhance intrinsic and extrinsic motivation, support usage of effective and flexible learning strategies, and facilitate students’ self-regulation (i.e., promoting flexible behaviors) resulting in better performance on achievement tasks (Pekrun et al., 2002). Positive deactivating emotions (relief or relaxation) distract students, reducing the cognitive resources for a task, resulting in poorer achievement. The effects of positive deactivating emotions on motivation are complex. For example, relaxation may have negative effects on intrinsic motivation but positive effects on extrinsic motivation if the external rewards in the future is to reach the state of relaxation again. Thus, the indirect effects of positive deactivating emotions on achievement via motivation may depend on the type of motivation. The effects of positive deactivating emotions on learning strategies and self-regulation and, in turn, on achievement is less understood but thought to be similar to those of positive activating emotions.

The effects of negative activating emotions (anger, anxiety) on achievement via cognitive resources and usage of learning strategies are thought to be negative but more complex for motivation and self-regulation. Anger and other negative activating emotions are thought to reduce cognitive resources by producing task-irrelevant thinking, resulting in negative effects on achievement. These emotions are also thought to negatively impact performance via use of more rigid learning strategies. The effects of negative activating emotions on achievement via motivation and self-regulation depend on further distinction. That is, by experiencing these emotions, intrinsic motivation may be reduced resulting in a negative impact on performance but extrinsic motivation could be induced in an effort to avoid future failure. Negative activating emotions tend to have differential
effects on self-regulation as well. On one hand, negative activating emotions may relate positively to external self-regulation (i.e., promoting reliance on teachers or instructors), thus increasing performance. Yet at the same time, negative activating emotions may relate negatively to internal self-regulation, thus decreasing performance. The effects of negative deactivating emotions (e.g., boredom) on performance via the cognitive resources, motivation, learning strategies and self-regulation are thought to be uniformly negative.

**Reciprocal causation in CV theory.** Thus far, a causal chain of variables has been described, in which environmental variables impact appraisals, which then influence emotions, which then affect achievement through cognitive-motivational variables (see Figure 2). Yet, according to CV theory, the effects of the constructs described above are reciprocal. For example, in the causal chain, CV theory specifies that emotions are related to motivation which then affects achievement. If achievement performance is not favorable, negative emotions may develop, which will then undermine future cognitive-motivational constructs and eventually, over time, will lead to poorer achievement.

Consider another example, where positive emotions foster positive subsequent situational appraisals, which will then cause subsequent positive emotions. That is, a student who experiences joy when doing homework, may see homework as an intrinsically rewarding activity, which will then foster positive emotions in the future.

The causal feedback loops between all constructs are present throughout the whole CV model (see Figure 2). Thus, CV theory should be evaluated using data collected over time, since the theory specifies that the variables change over time. In the following section, I summarize five studies, of varying design and analysis quality, that
attempted to investigate the reciprocal effects of emotions and achievement over time. The studies reviewed focus on how emotions affected subsequent achievement, and how achievement affected subsequent emotions. Thus, theoretically, the effects might be biased, since, according to CV theory, emotions affect achievement via variety of intervening variables such as motivation or cognitive resources (Pekrun, 2017). Moreover, as I note below, some studies employed a sequential design whereas others employed a longitudinal design. Using sequential designs researchers collect data over time, thus having the advantages of measuring constructs over time, but the disadvantages of lacking the control for previous levels of constructs. Longitudinal designs differ from sequential designs in that using longitudinal designs, all constructs are measured during each wave of measurement. Longitudinal designs have the advantage of controlling for previous levels of constructs when modeling estimating the effects of constructs involved in mediation. Thus, longitudinal designs provide a more rigorous test for effects involved in mediation.

Using a sequential design, Reeve, Bonacio, and Winford (2014) examined the effect of emotions on achievement and the subsequent effect of that achievement on emotions. In their study of college students, emotions were measured immediately before (joy, hope, pride, anger, shame and anxiety) and immediately after the exam (pride, relief, anger, and shame). They found positive (joy, hope, pride) and negative (anger, shame, and anxiety) test emotions measured before an exam related positively and negatively to exam scores, respectively. Exam scores had positive and negative effects on positive and negative retrospective test emotions that occurred after the exam. Although this research contributed to the study of test-taking emotions by examining the reciprocal
effect of test-taking emotions and test performance, several improvements could be made. First, as indicated above, CV theory explains how emotions affect subsequent performance, which then affect subsequent emotions. Using longitudinal design instead of the sequential design would provide stronger evidence about causal effects of emotions on subsequent performance. Furthermore, according to CV theory, the effects of emotions on test performance should be indirect via the cognitive-motivational variables, thus the model is misspecified.

Similar results were obtained by Putwain, Becker, Symes, and Pekrun (2018) when using a sequential design. They measured lesson-related boredom and enjoyment over four waves during a single school year. They found that lesson-related boredom affected subsequent mathematics achievement negatively in a sample of 5th and 6th graders. Achievement then had a negative effect on boredom later in the academic year. Additionally, a positive reciprocal effect of lesson-related enjoyment and achievement was also discovered. As with the Reeve et al. (2014) study, two major limitations exist: use of sequential research design and not including relevant variables that are hypothesized to mediate relationship between emotions and performance on a test.

Using a longitudinal design, the reciprocal effects of activity emotions and achievement were examined over five years for 5th-10th graders (Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017). Researchers measured positive and negative emotions at the end of the school year in each grade and then obtained end-of-year grades, which served as students’ achievement indication. They modeled the positive activity emotions of enjoyment and pride and the negative activity emotions of anger, anxiety, shame, and boredom. This study replicated the same pattern of relationships
between emotions and achievement as in the Reeve et al. (2014) study, in that positive emotions were positive predictors of subsequent exam scores, whereas negative emotions were negative predictors of subsequent exam scores. Exam scores predicted subsequent positive emotions positively and negative emotions negatively. This study serves as a replication of the Reeve et al. (2014) study with respect to findings positive and negative effects of emotions on subsequent task performance. Notably, this study used a longitudinal design, thus causal inferences from this study are more trustworthy; however, as Pekrun et al. (2017) admit, their study was not a true experiment. Additionally, lacking in this study is the inclusion of variables that are hypothesized to mediate the relationship between achievement emotions and performance. Thus, the parameter estimates for the effects of emotions obtained in this study are likely biased.

Using a longitudinal design, Pekrun, Hall, Goetz, and Perry (2014) found that for college students, class-related boredom had negative effect on subsequent achievement (measured by test scores), which then negatively affected subsequent boredom. Learning-related boredom was measured three times in the first semester, and two times over the second semester, each being measured at maximum 10 days after administering the course test. Similar to the Pekrun et al. (2017) study reviewed above, this study measured and modeled the relationship between boredom and performance longitudinally, but failed to include any of mediating variables specified by CV theory. Thus, the same limitations apply to this study as with Pekrun et al. (2017): parameter estimates associated with the reciprocal effects may be biased due to model mis-specification.

Using a longitudinal design over a year-long undergraduate chemistry course sequence, the positive reciprocal effects of enjoyment and achievement were replicated
when achievement was operationalized using chemistry exam scores (Gibbons, Xu, Vilafane, & Raker, 2018). Notably, all three types of emotions (class-related, learning-related, and test-related) were measured and modeled over time. However, a potential weakness in this study is that researchers modeled all three types of emotions at the same time. That is, enjoyment at time 1 was a composite of items measuring class-related, learning-related, and test-related enjoyment. Additionally, a negative reciprocal relationship was found for anxiety and test scores. A better approach to modeling the data would have been by separating the three types of emotions (i.e., class-related, learning-related, and test-related) and modeling the effects on exam scores simultaneously. However, three types of enjoyment (and anxiety) were highly correlated (i.e., $r$s ranging from .62 to .79), thus it likely that the composite was created to avoid issues associated with multicollinearity.

Studies investigating reciprocal feedback loops involving constructs other than emotions and achievement measures are lacking, hence the contribution of the current study given effort will be modeled. Using a sequential design, Putwain and colleagues (2018) investigated the reciprocal nature of learning-related emotions (i.e., enjoyment and boredom) and perceived control and value appraisals. Learning-related enjoyment and boredom, action-control expectancy (i.e., self-efficacy) and three types of subjective value (intrinsic, achievement, and utility) were measured over time. Specifically, researchers collected data at three times, each separated by three months. Data was collected sequentially; that is, enjoyment and boredom were measured at the first occasion, perceived control and values were measured at the second occasion, and enjoyment and boredom were measured again at the third occasion. Researchers
estimated six separate models (two emotions * three types of values). In the first model, enjoyment was measured at time 1, perceived control and intrinsic value were measured at time 2, and enjoyment was measured again at time 3. An interaction between perceived control and intrinsic value was included at time 2, and the effect of the interaction on enjoyment at time 3 was estimated. In the second model, enjoyment was measured at time 1, perceived control and achievement value were measured at time 2, and enjoyment was measured again at time 3. An interaction between perceived control and achievement value was included at time 2, and the effect of the interaction on enjoyment at time 3 was estimated. In the third model, enjoyment was measured at time 1, perceived control and utility value were measured at time 2, and enjoyment was measured again at time 3. An interaction between perceived control and utility value was included at time 2, and the effect of the interaction on enjoyment at time 3 was estimated. The next three models included boredom, instead of enjoyment and the variations of models remained the same. Results from all six models showed that enjoyment and boredom had effects on subsequent perceived level of control and each type of subjective control. For enjoyment these effects were positive, whereas for boredom these effects were negative. The effects of the situational appraisals (perceived control, subjective values, and the interactions between them) were mixed. That is, two significant interactions were observed. There was a significant positive interaction between perceived control and achievement value when modeled to predict subsequent enjoyment. That is, higher levels of interaction effect of perceived control and achievement value was related to higher levels of students’ enjoyment. There was a significant negative interaction between perceived control and intrinsic value when modeled to predict subsequent boredom. That is, higher
levels of interaction effect of perceived control and intrinsic value was related to lower levels of students’ boredom. The remaining four models revealed main effects of either perceived control or a type of subjective value on subsequent enjoyment. Intrinsic value in one model, and utility value, along with perceived control in another model had positive effects on subsequent enjoyment. None of these effects were statistically significant in predicting subsequent boredom.

This study by Putwain and colleagues (2018) is important to the literature of achievement emotions because to my knowledge, it is the first study to include the interaction effect of perceived control and type of subjective value on achievement emotions. Furthermore, researchers included achievement value, which has not been examined before, and interactions between perceived value and type of subjective value. Results are intriguing; however, several limitations exist. First, the use of sequential research design limits inferences. As mentioned multiple times before, longitudinal research design with autoregressive effects should be utilized to obtain accurate parameter estimates. Secondly, researchers noted that multicollinearity between class-related, learning-related, and test-related emotions prevented them from modeling all emotions at the same time. Thus, they were unable to examine the effects of one type of value on emotions when controlling for other types of values. For example, it is unclear if intrinsic value would have an effect on enjoyment, after controlling for achievement value and utility value.

Currently, studies examining the complete CV reciprocal model are non-existent. Researchers have only recently started to investigate the effects of emotions on subsequent achievement and the effects of achievement on subsequent emotions.
(Gibbons et al., 2018; Pekrun et al., 2014; Pekrun et al., 2017; Putwain et al., 2018; Reeve et al., 2014). Five of those studies were reviewed here. Although these studies are encouraging, as they provide evidence for the reciprocal effects in CV theory, the effects of these studies should be carefully evaluated given design and analysis issues. It is worth repeating that according to CV theory, the effect of emotions on subsequent achievement are hypothesized to be indirect via motivation, cognitive resources, and learning strategies. Only one study examined the effects of emotions on subsequent situational appraisals and the effects of situational appraisals on subsequent emotions. Moreover, to my knowledge, no studies have investigated the feedback loop between emotions and motivation. Hence, the current study examines the reciprocal effects of appraisals, emotions, and motivation during the achievement activity of completing a test. Importantly, a longitudinal design was employed to estimate the theorized mediated effect of appraisals on effort via emotions.

Mediation

Given the mediation hypotheses specified in CV theory, a thorough understanding of the concept of mediation, the conceptual and statistical assumptions, and the necessary data and research design to statistically evaluate the mediation mechanism is required. First, I define mediation. Second, I describe and critique three research designs used to gather data for mediation analyses: cross-sectional, half-longitudinal, and fully longitudinal. I use examples from the current study to illustrate the differences between these research designs and the inferences that can be made from using them. More specifically, I explain three concepts of change (i.e., stability, stationarity, and equilibrium) and how they relate to different research designs used for mediation.
Mediation is the “way in which one variable (generically, X) has an effect on another variable (Y) through its influence on some intermediate variable (M)” (Selig & Preacher, 2009, p. 144). In other words, mediation depicts the process or the mechanism that explains the causal effect of one variable on another, via third variable(s). Describing mediation requires certain terminology. In mediation analysis, inferences about the effect of one variable on another are desired, rather than simply stating that variables are related. That is, causal language is used to describe mediation. A common approach to illustrate mediation is via path diagrams (see Figure 3). Referring to Figure 3, the space between X and M, M and Y, and X and Y is used to show there is (or should be) a time lag between variables, as temporal precedence is required for the causal effect to be inferred.

Before discussing mediation, consider a simple regression model (Figure 3a). In this model, X is measured at time 1 and Y is measured at time 2; thus, there is a time lag between measures of both variables. Assume that the relationship (measured by correlation coefficient) between X and Y is known and is .5. The path diagram illustrates that X has an effect (c) on Y. The direct effect can be expressed as:

\[ Y = i + cX + e \]

Where i refers to an intercept, c is the prediction coefficient, and e is the error. The simple linear regression model represents a theory, which states that X directly causes Y. The effect of X on Y is depicted as “c” path in the figure. The standardized c effect would be equal to .5, which is equal to the observed relationship between X and Y. In other words, the effect of X on Y completely represents the observed relationship between the two variables.
The mediated effect of X on Y can be complete or partial (Figure 3). To establish complete mediation, a few conditions must be met. First, X must have a direct effect to M (the mediator). In Figure 3c coefficient $a$ represents the direct effect of X on M. Second, M must have a direct effect on Y (path $b$). The direct effects can be expressed by two equations:

\[ Y = i + bM + e \]
\[ M = i + aX + e \]

If there is no direct effect of X on Y, then M completely mediates the relationship between X and Y via an indirect effect. The indirect effect is simply the product of $a$ and $b$ path coefficients. In other words, the relationship (i.e., correlation) between X on Y is hypothesized to be completely explained via variable M. Consider the hypothetical example again, in which the observed relationship between X and Y is equal to .5. If the complete mediation model is correct, then the direct effect of X on Y is equal to 0; however, the model-implied relationship would equal .5. The effect of X on Y is mediated by M. Using the two formulae above, the effect of X on M and the effect of M on Y can be computed. If the correlation of X and M equaled .71, then the effect of X on M ($a$) would equal .71. If the correlation of M and Y equaled .71, then the effect of M on Y ($b$) would equal .71. The indirect effect of X on Y via M would equal .71 * .71 or .5. The mediation model reproduces the observed relationship between X and Y (.50) perfectly. With an appropriate research design, such a result would provide evidence that complete mediation model (as depicted in Figure 3c) is feasible.

If X has a direct effect ($c'$) on Y (Figure 3b), in addition to the indirect effect of X on Y via M, then the relationship between X and Y is said to be partially mediated. That
is, the product of relationship between X and M and M and Y does not sufficiently explain the observed relationship between X and Y. The observed relationship between X and Y is explained by two effects: the direct effect of X on Y and the indirect effect of X on Y via M. The formulae to illustrate partial mediation model are:

\[ Y = i + c'X + bM + e \]
\[ M = i + aX + e \]

The equation used to predict Y now includes \( c' \) term that represents a direct effect of X on Y. Thus, to calculate the total effect of X on Y, the direct effect (\( c' \)) should be added to the indirect effect (\( a*b \)). Partial mediation models are just-identified when modeling three variables, which means the model will reproduce the observed relationships perfectly. Parameters (\( a, b, c' \)) in partial mediation models can be estimated; however, the model cannot be tested for fit. Assume that the observed relationship between X and Y is equal to .5. To reproduce this XY relationship, let \( a \) and \( b \) equal to .6 and \( c' \) equal to .14. As demonstrated above, to calculate the indirect effect of X on Y via M, a product of two effects should be obtained. Using this example, the indirect effect would be \( .6 * .6 = .36 \). The total effect is calculated in partial mediation model by adding the indirect effect (\( a*b; \) via M in this example) and the direct effects of X on Y (\( c' \)). Thus, the total effect is \( .36 + .14 = .5 \), which matches the observed relationship between X and Y.

In the current study, I specified a longitudinal panel model that involves perceived control (i.e., self-efficacy), value (i.e., achievement value), emotions, and motivation (i.e., effort). This panel model includes several tests of complete and partial mediation. As part of the larger panel model, I will investigate if test-taking emotions (M) are mediating the effects of test-taking importance (X) on test-taking effort (Y). If I had a non-mediated
(simple linear regression) hypothesis, then I would be interested in how importance (X) directly affects test-taking effort (Y). However, according to CV theory, the effect of importance (X) on test-taking effort (Y) is mediated by test-taking emotions (M). To investigate complete mediation, I would be interested if the effect of importance on effort is completely explained by the test emotions. To investigate partial mediation, I would be interested if, in addition to the indirect effects (i.e., the effect of importance on motivation via test emotions), the direct effect of importance on motivation exists.

Importantly, given the hypothesized effects of appraisals (e.g., perceived control and value), emotions (i.e., anger, joy, boredom, and pride), and motivation (e.g., effort) as specified by CV theory, longitudinal data with autoregressive effects is necessary to appropriately estimate the effects of mediation. In the next section, I review different research designs and the inferences about mediation that can be drawn from using each of the research designs.

**Research design and inferences.** Next, I review the most popular ways to obtain data for mediation analysis. There are four widely used research designs: cross-sectional, sequential, half-longitudinal, and fully-longitudinal. For each, I describe the typical data collection process, models tested, and inferences that can be made regarding mediation process.

In this section, I also describe three concepts of change (i.e., stability, stationarity, equilibrium) that are relevant to discussion of mediation and link it back to each research design. In short, the main point emphasized below is that modeling mediation from cross-sectional or sequential data is not appropriate if the goal is to infer causality. Even though such practice is commonly used in social sciences, researchers have explained that a
cross-sectional research design ignores the role of time, which is a critical limitation to properly modeling mediation (Preacher, 2015). The mediation process, by definition, specifies the need for longitudinal research designs that allow time lags between variables (Cole & Maxwell, 2003; Jose, 2016). Longitudinal data improves the strength of argument for causality by measuring variables over time, thus change can be evaluated. Thus, the limited inferences afforded by of cross-sectional research design will become apparent after comparison to the more accurate inferences afforded by the uses of longitudinal research designs in the next section.

**Cross-sectional design.** Mediation processes take time to show effects (Cole & Maxwell, 2003; Jose, 2016). For example, test-taking importance is hypothesized to cause subsequent emotions, which then affects subsequent motivation (Pekrun, 2006, 2007, 2017). When data is collected concurrently using cross-sectional designs (Figure 4a), there is no time lag between measuring each variable, which presents a logical error in inferring causation. Using cross-sectional data to estimate mediation, the causal hypothesis would be that variable X immediately causes variable M and variable M immediately causes variable Y. Such a theory is unlikely to be true. Rarely are constructs are theorized to affect other constructs instantaneously.

The majority of mediation hypotheses do not theorize that constructs affect subsequent constructs immediately; thus, cross-sectional data severely limits the inferences made about mediation mechanisms. To illustrate the disadvantage of using cross-sectional data, keep in mind that mediation analysis is used for testing causal relationships between variables. Causality implies that change in one variable (X) causes
a change in another variable (Y). There are three different concepts of change: stability, stationarity, and equilibrium.

**Stability.** Stability is simply the opposite of change (Cole & Maxwell, 2003; Little, 2013; Mitchell & Maxwell, 2013; Preacher, 2015). Some variables do not exhibit change in magnitude or rank order of individual scores over time (i.e., they are constant over time). Thus, the term “stability” is used to refer to unchanging mean levels of a construct over time. For example, empirical studies have found the average of test-taking importance tends to be stable throughout a low-stakes test (Finney, Satkus, & Perkins, 2018). Stability also refers to the extent to which individual differences in the variable are maintained over time (i.e., unchanging rank order of subjects over time; Preacher, 2015). For example, consider a hypothetical situation in which test-taking importance was measured twice: before a test (time 1) and at the end of the test (time 2). A correlation coefficient between time 1 and time 2 could be calculated. If the coefficient was close to 1, such finding would imply that individuals are not changing their rank order on test-taking importance over time. In other words, the rank order of individuals is stable across time.

**Stationarity.** Stationarity refers to the unchanging causal structure of variables over time (Cole & Maxwell, 2003; Little, 2013; Mitchell & Maxwell, 2013; Preacher, 2015). Consider a scenario in which test-taking importance, test-taking anger, and test-taking effort were measured over five waves of measurement (Figure 5). To examine stationarity of the direct effect of test-taking anger on test-taking effort, researchers would have to estimate the directional effects of test-taking anger on test-taking effort on multiple occasion. That is, the direct effect of test-taking anger at time 1 on test-taking
effort at time 2 should be compared with the direct effect of test-taking anger at time 2 on test-taking effort at time 3. Additionally, the direct effect of test-taking anger on test-taking effort could be estimated at time 3 to time 4 and time 4 to time 5. If all of these effects (i.e., $AE_{12}$, $AE_{23}$, $AE_{34}$, $AE_{45}$) were equal in magnitude, stationarity would be established. In other words, stationarity would mean that during every measurement of test-taking anger and test-taking effort, the same causal structure is in place – test-taking anger causes subsequent test-taking effort (depicted as a direct effect $AE_{xx}$ in Figure 5).

When examining mediation processes, the indirect effect is the focus of the analysis. To evaluate stationarity for the indirect effect of importance on test-taking effort via anger, there are several effects that would be estimated and compared. The indirect effect of importance at time 1 on effort at time 3 via anger at time 2 could be estimated ($IA_{12} \times AE_{23}$). The indirect effect of importance at time 2 on effort at time 4 via anger at time 3 ($IA_{23} \times AE_{34}$) could be estimated. Lastly, the effect of importance at time 3 on effort at time 5 via anger at time 4 ($IA_{34} \times AE_{45}$) could be estimated. If all of these indirect effects were of equal magnitude, then strong evidence of stationarity of this indirect effect would be provided. Some theorists assume the causal process reflected in a mediation model does not change over time (i.e., stationarity is established). Much like in the example of the direct effect, if stationarity of the indirect effect is established, then the mediated effect (how changes in X lead to changes in Y via M) would be the same whenever variables of interest are measured.

A finding of stationarity is important for practitioners because it can inform interventions by excluding a potential confound in timing of the causal process. In other words, if stationarity is established, then the timing of when the intervention (X) is
delivered would not make a difference. However, if stationarity of indirect effect is not established, such finding would suggest that the relationships between variables involved in mediation are complex and the effects between variables will depend on when the variables are measured. For practitioners, lack of stationarity means that careful consideration of time points is warranted because collecting data at different time points may result in different indirect effects. In other words, mediation effects may be moderated by the time occasion.

*Equilibrium.* Equilibrium is described as stable cross-sectional variances and covariances over time (Cole & Maxwell, 2003; Little, 2013; Mitchell & Maxwell, 2013; Preacher, 2015). Equilibrium can be tested by inspecting within-wave (cross-sectional) variances and covariances and evaluating if they are equal over time. Cole and Maxwell (2003) explain that, over time, a model that exhibits stationarity will reach equilibrium; however, this may take several waves. The concept of equilibrium is related to the assumption of sphericity that is made in traditional statistical analysis of repeated measures ANOVA. A model or a system of variables that does not exhibit equilibrium might suggest that it is misspecified and further investigation should be warranted.

To illustrate how to test for equilibrium, recall a scenario in which importance, anger, and effort were measured five times (Figure 5). To test for equilibrium, the variances of importance, anger, and effort measured at time 1 should be compared to the variances of importance, anger, and effort measured at time 2, time 3 and so on. Additionally, the covariances between the three variables at time 1 should be compared to the covariances between the three variables at time 2, time 3 and so on. If, say, the variance of effort at time 5 is different from the variances of effort at time 3 and time 4,
then this result would imply that some other unknown (e.g., unmeasured) variable may be affecting the variability of effort scores at time 5 and the system of variables is not at equilibrium. In other words, apart from the already specified effects of anger at time 4 and effort at time 4 on effort at time 5, a new, perhaps unmeasured variable is affecting effort only at time 5.

In the context of test-taking, perhaps, examinee’s fatigue at time 5 becomes important and reduces the variability of effort scores. Such a hypothesis is plausible, but impossible to test using the current model (Figure 5) because the mediation model does not include fatigue that exerts its influence on effort only at time 5. In other words, the model is misspecified since it does capture all causal processes occurring over time.

Failing to achieve equilibrium is not a disappointing result. Jose (2016) reflects that equilibrium will rarely be established in social and behavioral sciences, and so researchers should not be discouraged if this happens in their studies. However, failing to establish equilibrium weakens the inferences that can be made in mediation. That is, if equilibrium is established and mediation is observed, then one could draw conclusions that the system of variables involved in mediation model is stable and over time will retain the effects (effect of X on Y, Y on M and X on Y via M) involved in mediation. In other words, by establishing equilibrium researchers can be more confident that the mediation models estimated will replicate in the future contexts and different samples. In contrast, if equilibrium has not been achieved then practitioners should review their theory and investigate what other constructs may be causing the system of constructs to be unstable. Mediation can still be inferred, however, without equilibrium it would be unclear whether a particular mediation model would replicate in the future.
Why are these concepts important to mediation? Stability, stationarity, and equilibrium are all related to change in a different way, which is exactly what mediation analysis examines. Stability represents no change in individual differences on a variable over time. Stationarity represents no change in causal structure of variables over time. Equilibrium represents no change in variances and covariances over time. When modeling cross-sectional data, stability, stationarity, and equilibrium cannot be tested because there is a lack of longitudinal data. When using cross-sectional data there is no way to empirically test for stability, stationarity and equilibrium; thus, these concepts have to be assumed. Making such assumptions weakens the inferences one can make about mediation. For example, without testing stability of mediation effects, one would have to assume there are no individual differences on variables involved in mediation over time, which is a rather strong assumption. If stationarity is not empirically tested, then one would have to assume the estimated mediation effects would be of equal magnitude over time. Lastly, having to assume equilibrium would imply that during each new wave of measurement variances and covariances of all variables are equal to subsequent variances and covariances. Thus, when using a cross-sectional research design, conclusions about mediation are weak. Researchers should not make claims about the directionality of the effects of one variable on another since such claims necessitate time lags between measures of each variable.

**Sequential design.** The major weakness of cross-sectional research design is the lack of time lag between the same variables involved in mediation. Using sequential design, variables are measured over time (Figure 4b). That is, variable X is be measured first, then after some time variable M is measured, and then after some time variable Y is
measured. Similarly, to cross-sectional design, three variables are measured once. In contrast to a cross-sectional research design, using a sequential design, variables are measured at three different measurement occasions. Thus, the requirement of having time lag between measures is satisfied. However, methodologists have provided empirical evidence that mediation effects estimated using sequential data are biased in the same way as the mediation effects obtained by cross-sectional data (Mitchell & Maxwell, 2013). Time lags between measures afforded by sequential design are not sufficient to infer mediation. Additionally, Mitchell and Maxwell (2013) recommend that prior levels of each variable should be measured and controlled for to obtain accurate estimates of direct effects.

In regard to the concepts of change outlined above, the sequential design does not offer any improvement over cross-sectional data. In fact, stability, stationarity, and equilibrium cannot be tested using the sequential design because as with cross-sectional data, each variable is measured only once. Thus, researchers have to assume these concepts exist, which weakens the argument for mediated effects.

**Half-longitudinal design.** The difference between a sequential design and a half-longitudinal design is that using sequential design, all variables are measured at each wave of measurement (Figure 4b and 4c respectively). For example, using a sequential design, only variable X would be measured at time 1, only variable M would be measured at time 2, and only variable Y would be measured at time 3. Using a half-longitudinal design, the three variables (X, M, and Y) would be measured at two different times.
An example with variables from the present study is shown in Figure 4c. There are two waves of concurrent data measured with some theoretically appropriate time-lag between them. This design allows estimation of the effect of importance at time 1 on anger at time 2 (IA₁₂) or estimation of the effect of anger at time 1 on effort at time 2 (AE₁₂). These effects are called *cross-lagged effects* because these effects refer to how variable X (e.g., importance) at time 1 affects variable M (e.g., anger) at time 2. The term “cross-lagged” means the effect of variable X on variable M is estimated across a time lag. These cross-lagged effects are interpreted as direct effects of variable X at an earlier time-point on variable Y at a later time-point, after controlling for other variables in the model.

An indirect effect of variable X on variable Y via variable M would be computed as the product of paths that are leading from the variable X to variable Y via other variables and interpreted as how variable X is exerting its influence on Y via other variables. To illustrate consider Figure 4c. In a half-longitudinal design, an example of an indirect effect would be the product of paths IA₁₁*AE₁₂ and IA₁₂*AE₂₂.

Lastly, the sum of the direct effect and all possible indirect effects is called the total effect and is interpreted as how variable X is influencing variable Y through all possible pathways. For example (Figure 4c), in order to estimate the total effect of importance at time 1 on effort at time 2 all possible paths should be added together. Notice, there is no direct effect of importance at time 1 on effort at time 2. In other words, according to this model, the direct effect is set to 0. However, there are two indirect routes that relate importance at time 1 to effort at time 2. To calculate the total effect, the effect of importance at time 1 on effort at time 2 via anger at time 1
(IA_{11} \times AE_{12}) should be added to the effect of importance at time 1 on effort at time 2 via anger at time 2 (IA_{12} \times AE_{22}).

In addition to the cross-lagged effects, a half-longitudinal research design allows us to estimate autoregressive effects. An example of an autoregressive effect is the effect of importance at time 1 on importance at time 2 (II_{12}), or any construct from time 1 to the same construct at time 2. Autoregressive effects are interpreted as a level of stability between the same construct across two time-points. In other words, autoregressive effects (sometimes called stability coefficient) indicate how stable a construct is over time.

Half-longitudinal research designs are more appropriate to model mediation than cross-sectional designs for two reasons. First, modeling mediation using half-longitudinal design is more aligned with theory, which suggests that some time should elapse between the constructs of interest, so that variable X can cause variable M (or M cause Y). Second, the ability to estimate autoregressive effects in half-longitudinal designs allows causal language. Autoregressive effects in a simple three-variable context could be thought of as a semi-partial correlation coefficient that takes into account the effects of other variables in the model. If the autoregressive effect is equal to one, that means that from time 1 to time 2, the rank order of individuals has not changed, which would imply the influence of other variables will be equal to 0. If the autoregressive effect is lower than a value of one, subjects have changed rank order, and effects of other variables on variable M (or Y) could explain why individuals changed the rank order. Usually, individuals tend to not move up or down in rank order on a particular construct over a short amount of time, especially if measured using the same instruments (Cole & Maxwell, 2003). High autoregressive effects are important to consider when talking
about cross-lagged effects because the cross-lagged effects may appear to be small in magnitude, but are important nonetheless.

Recall, using cross-sectional or sequential data, the concepts of change in mediation has to be assumed. Half-longitudinal design offers an improvement in this area. Regarding stability, with two time-points, the mean levels of each construct could be compared and tested for equality across time. Similarly, the rank order of individuals between measures of the same constructs over time could be estimated via the autoregressive coefficient. High autoregressive coefficients would provide some evidence that rank order of individuals on variable X is stable across time (i.e., from time 1 to time 2). Thus, with half-longitudinal design, stability can be evaluated and does not need to be assumed (as with cross-sectional data and sequential), which strengthens the argument made for mediation.

Regarding equilibrium, the variances and covariances of variables within time-points can be compared across time-points. Recall from the discussion above that the finding that the variances and covariances are equal across time points (equilibrium is established) would suggest the model is correctly specified and there are no additional important variables to consider at different times. The inequality of variances and covariances would mean the opposite result; the system of variables is not stable, some constructs are more important to consider at different time points than others. Using a half-longitudinal design, which consists of two waves of measurement, equilibrium can be tested, which provides an improvement to the previous two research designs.

Regarding stationarity, the unchanging causal structure of interest in mediation analysis is the indirect effect of X on Y via M or in Figure 4c, the indirect effect of
importance on effort via anger. The major limitation for half-longitudinal design is inability to have time lags between both cross-lagged effects (i.e., effect of importance at 1 time on anger at time 2 or the effect of anger at time 1 on effort at time 2) involved in a mediation model. Without the ability to properly test for the indirect effect, stationarity cannot be established. That is, using half-longitudinal design it is impossible to not only test if the indirect effect would be stable (stationarity) but to appropriately test the indirect effect in the first place. In summary, the half-longitudinal design provides an upgrade over a cross-sectional design in that stability and equilibrium can be empirically evaluated and established. However, stationarity cannot be tested, which means that different design should be used.

In summary, the half-longitudinal research design is an improvement from the cross-sectional and sequential data designs because it allows to have some time between the variables and we can control for the previous levels of constructs via autoregressive effects. However, the limitation of half-longitudinal design is that the indirect effect of X on Y via M cannot be estimated with appropriate time lags between each effect (the effect of X on M and the effect of M on Y).

**Fully-longitudinal design.** Consider the three waves of measurement in Figure 4d, which illustrates fully longitudinal research design. The major improvement from the half-longitudinal research design is the additional wave of measurement. This design gives the ability to estimate mediation, and more specifically, the indirect effect of X on Y via M, as theorized by mediation hypothesis. Using variables from this study, the cross-lagged effect of importance at time 1 on anger at time 2 (IA_{12}) and the cross-lagged effect of anger at time 2 on effort at time 3 (AE_{23}) can be estimated. The two cross-lagged
effects are estimated when taking into account the autoregressive effect of anger at time 1 on anger at time 2 and the autoregressive effect of effort at time 2 on effort at time 3.

The longitudinal research design provides an additional wave of measurement which can be used to test for stability. For example, the rank order for individuals at time 1 for variable X (e.g., importance at time 1 in Figure 4d) can be compared with the rank order for individuals at time 2 for variable X (e.g., importance at time 2 in Figure 4d), which can be compared to the rank order for individuals at time 3 for variable X (e.g., importance at time 3 in Figure 4d). Recall, evaluating stability is possible with half-longitudinal design; however, using a fully-longitudinal research design, the test for stability is more rigorous. With additional waves of measurement in a fully-longitudinal design, there will be more ways that stability can be evaluated. Naturally, the more ways stability can be tested, the more ways it can fail to be established. However, the inference about stability from a research design that has established stability with five waves of measurement is more trustworthy than inference about stability from research design with three waves of measurement. To sum up, although it is possible to test stability with a half-longitudinal design, the results from fully-longitudinal design will be more trustworthy, especially with greater number of waves of measurement.

With three waves of measurement, stationarity can only be partially tested. Recall, stationarity refers to unchanging nature of causal process. With three waves of measurement, a comparison of the cross-lagged effect from time 1 to time 2 and time 2 to time 3 is possible. I say partially because there still is only one way to estimate the indirect effect (Figure 4d: IA12 * AE23). Thus, stationarity is still being assumed with only 3 waves of measurement as there is no evidence if the pattern of indirect effect will
remain same in terms of magnitude over time. However, with 5 waves of measurement (see Figure 5), three potential indirect effects can be tested. That is, the effect of importance at time 1 on effort at time 3 via anger at time 2 (IA$_{12}$ * AE$_{23}$) can be estimated. Next, the effect of importance at time 2 on effort at time 4 via anger at time 3 (IA$_{23}$ * AE$_{34}$) can be estimated. Third, the effect of importance at time 3 on effort at time 5 via anger at time 4 (IA$_{34}$ * AE$_{45}$) can be estimated. Using this example, stationarity for the indirect effect can tested by comparing the first indirect effect (IA$_{12}$ * AE$_{23}$) with second indirect effect (IA$_{23}$ * AE$_{34}$), and comparing the second indirect effect (IA$_{23}$ * AE$_{34}$) with the third possible indirect effect (IA$_{34}$ * AE$_{45}$). This illustration suggests that to test stationarity, a fully-longitudinal design with at least four waves of measurement is necessary.

Lastly, a fully-longitudinal design does improve the ability to test equilibrium by adding an additional wave of measurement. Similarly, as with stability, to test for equilibrium, at least two waves of measurement are necessary. Two waves of measurement could be obtained with half-longitudinal design. However, by having three or more waves of measurement, the test for equilibrium becomes more rigorous, similar to the advantages of testing stability in longitudinal design versus half-longitudinal design. With three waves of measurement (Figure 4d), a simple way to test for equilibrium would be to compare the variances of variables at time 1 (i.e., importance, anger, effort) to the variances of the same variables at time 2 and time 3. The same procedure could be repeated for the covariances among the variables at time 1 to the covariances between the same variables at time 2 and time 3. Thus, the inferences about mediation would be stronger using fully-longitudinal design than half-longitudinal design.
because to establish equilibrium more equalities should be established. Table 6 summarizes each research design and highlights what concepts of change have to be assumed and what concepts can be tested.
Chapter 3: Method

Participants and Procedure

Data were collected during a large-scale institutional accountability testing session (“Assessment Day”) at a mid-size public university in the eastern United States. Students at this institution complete mandatory testing twice during their college careers: first, in the summer, the weekend before they begin their first-year at the university; second, in the Spring semester, after they accumulate between 45 and 70 credit hours (usually in the second or third year of their studies). The test scores from the two Assessment Days are used to evaluate student learning and development and are reported for state and regional institutional accountability purposes. The testing sessions are approximately two hours in length. Students complete both cognitive and non-cognitive tests.

There are no personal consequences for students for poor test performance on tests completed during Assessment Day. Thus, the tests are low stakes for students. An instructional video and trained proctors explain (via standardized instructions) that test results are used by university’s faculty and administrators to improve students’ academic and co-curricular experience.

Students who do not attend Assessment Day receive a hold on their student account, which prevents them from registering for classes in the subsequent semester. These students are invited to attend a make-up session to remove the hold. After students attend the make-up session, the hold from their account is released.

Data for the current study were collected during Assessment Day and during make-up testing sessions in Fall 2018. For Assessment Day, all incoming students were
randomly assigned (using the last three digits of student identification number) to a specific testing room, where a particular configuration of tests was administered. For the current study, data were collected from two test configurations. The first test configuration included two tests administered in the following order: a Quantitative and Scientific Reasoning Test (NW9) and an Information Literacy Test (INFOCORE2). The second test configuration included two tests administered in the following order: a Quantitative and Scientific Reasoning Test (NW9) and an Oral Communication Test (OCP3). Data from the second test (either INFOCORE2 or OCP3) were not considered in this study and were mentioned to characterize the full two-hour testing session. The NW9 served as the performance measure and students completed measures of self-efficacy, perceived value, examinee effort, and test emotions during this test to evaluate hypotheses stated below. During Assessment Day, data were collected across two test platforms: paper and pencil and computer.

Students who did not attend Assessment day were invited to attend a make-up session. Three make-up session were held in the first three weeks of the Fall 2018 semester. In the make-up sessions, students completed two tests in the following order: a Quantitative and Scientific Reasoning Test (NW9) and a Socio-Cultural Awareness Test (SDA7). Data from the SDA7 were not considered in this study. Data from the make-up sessions were collected via computer. A total of 179, 18, and 48 students attended the first, second, and third make-up sessions, respectively. Thus, data from 245 students who attended the make-up sessions were considered in this study. In sum, 749 students from the Assessment Day and 245 students from the make-up sessions were combined to form the initial sample of 994 students.
After inspecting the data, 39 students were removed due to incomplete data. Thus, the effective sample size is 955. The demographics of Assessment Day and make-up sessions samples (i.e., gender, age, and ethnicity) are provided Table 5. The demographics approximate the university’s population of incoming first-year students. More male than female students attended make-up sessions (51.25 % vs. 48.75 %), which is common (Barry & Finney, 2016; Brown & Finney, 2011; Swerdzewski, Harmes, & Finney, 2009). Age in years was approximately equal between the two samples (18.45 vs 18.50 for assessment day and make-up sessions, respectively).

**Measures**

*Quantitative and scientific reasoning.* Quantitative and scientific reasoning ability was assessed using the Natural World Test, version 9 (NW9, Sundre & Thelk, 2010). The NW9 served as the test to collect self-efficacy, test emotions, perceived importance, and expended effort. The NW9 is a 66 multiple-choice item instrument developed by faculty in the university’s general education program. Items on the NW9 are scored dichotomously as correct or incorrect. The test is difficult; first-year students answer approximately 70% of items correctly (Barry & Finney, 2015). The NW9 takes approximately an hour to complete.

Before starting the NW9, students were provided an example item representative of the content and difficulty of items on the NW9 (see Appendix A). Students had to be exposed to an example item to formulate self-efficacy perceptions. That is, without knowing what the test entails, students could not accurately evaluate their expectations of success on the NW9. After examining the example item, students completed measures of
self-efficacy and perceived test importance (described below). After completing these measures, students began the NW9.

To model expectancy, value, emotions, effort, and performance longitudinally, the NW9 was broken into three subtests, each containing 22 NW9 items. After completing subtest 1 (first 22 items on NW9), students rated their perceived test importance, expended effort, and levels of five test emotions (described below). Test instructions directed students to indicate their “feelings at that moment, after the first part of the test”. Next, students completed subtest 2 (next 22 items on NW9) and measures of perceived test importance, expended effort, and five test emotions. Lastly, students completed subtest 3 (last 22 items on NW9) and measures of perceived test importance, expended effort, and five test emotions. In sum, students completed a measure of self-efficacy once (before the start of the test). Students completed a measure of perceived test importance four times (once before the test, after subtest 1, after subtest 2, after subtest 3). Students completed measures of examinee effort and test emotions three times (after subtest 1, after subtest 2, after subtest 3). In the current study, internal consistency reliability was adequate for subtests 2 and 3 but not adequate for subtest 1 (see Table 6). These results inform the models estimated in the current study, as explained below in the Data Analysis section.

The NW9 was chosen as the test to examine the effects of self-efficacy, test importance, and test-emotions on examinee effort for two reasons. First, the regional and state accountability bodies that accredit the university require evidence of student learning in several areas. Quantitative and scientific reasoning is one of those areas. Thus, providing evidence of score validity for NW9 is pertinent to the university. Hence,
understanding how emotions and motivation is experienced during this test can inform inferences from the NW9 scores. Second, the subject of mathematics is often accompanied with a variety of emotions. Accordingly, in many empirical studies framed using CV theory, student emotions referring to the subject of mathematics were measured (Frenzel, Pekrun, & Goetz, 2007; Pekrun et al., 2017). To do so, researchers often adapt the Achievement Emotions Questionnaire (AEQ) to measure emotions experienced with respect to activities and outcomes in mathematics domain (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011).

Test emotions. Anger, joy, pride, and boredom were measured using items adapted from the Test-Emotions Questionnaire (TEQ; Pekrun, Goetz, Perry, Kramer, Hochstadt, & Molfenter, 2004) and items from a study investigating single-item indicators of state test emotions (Goetz, Preckel, Pekrun, & Hall, 2007). The TEQ was developed by adapting the AEQ, which measures emotions in achievement settings (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). That is, the TEQ items measure test emotions, whereas the AEQ items measure learning-related or course-related emotions. Anger, joy, pride, and boredom were measured using three items each. Worry was measured using three items that were adapted from Worry and Emotionality Scale (WES; Liebert & Morris, 1967) and the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1991).

Students responded to test emotions items using a 1 (“Strongly Disagree”) to 5 (“Strongly Agree”) Likert scale (see Appendix B for items and instructions). Higher scores on the items indicate higher levels of each test emotion. The instructions directed students to indicate their feelings after the first, second, and third part of the test. In other
words, state emotions experienced during the test were measured. In the current study, internal consistency reliability was adequate for all emotion scores at all measurement occasions (see Table 6).

**Perceived test importance and examinee effort.** Perceived test importance and expended effort were measured using the two subscales of the Student Opinion Scale (SOS; Thelk, Sundre, Horst, & Finney, 2009). The SOS is a 10-item self-report instrument with five items measuring the perceived value of the test and five items measuring expended examinee effort. Students respond to items using a 1 (“Strongly Disagree”) to 5 (“Strongly Agree”) Likert scale. There are four negatively-worded items (two for the perceived value subscale and two for the expended effort subscale) that need to be reverse-coded prior to computing subscores, which can range from 5 to 25. Higher scores indicate higher levels of each construct. Adequate internal consistency measures (Cronbach’s coefficient α) for each subscale have been obtained in the past studies (Thelk et al., 2009; Finney, Mathers, & Myers, 2016).

The SOS was originally developed to measure students’ importance and motivation after a testing session (Thelk et al., 2009). More recently, the SOS has been used to measure students’ importance and motivation for a specific test rather than a testing session (Finney, Mathers, & Myers, 2016; Mathers, Finney, & Hathcoat, 2018). The difference between the session-specific SOS and the test-specific SOS is the context that students rate their perceived importance and expended effort. The two-factor structure of the test-specific SOS was demonstrated in previous studies (Finney et al., 2016) and aligns with the two-factor structure obtained for the session-specific SOS.
The SOS properties have been shown to be invariant across computer-based and paper-and-pencil testing mediums (Thelk et al., 2009).

As explained above, students completed the test-specific SOS three times throughout the NW9. Thus, minor modifications to the wording of the SOS (e.g., adjusting the object focus from the test to the part of the test) and test instructions were necessary (see Appendix A). In the current study, internal consistency reliability estimates were adequate for expended effort scores at all measurement occasions (see Table 6). Internal consistency reliability was not adequate for perceived value before the test ($\alpha = .68$); thus, perceived value at time 0 was not modeled in the study (as explained below).

**Self-efficacy.** The Questionnaire on Current Motivation (QCM, Vollmeyer & Rheinberg, 2006) has been used to measure expectancy in previous studies of motivation in low-stakes testing contexts (Penk & Richter, 2017; Penk & Schipolowski, 2015). Thus, three items adapted from the QCM were used to measure students’ self-efficacy to succeed on the test (see Appendix A). Specifically, three items from the Probability of Success subscale were used, as they reflect students’ expectancy to succeed on a test. Students responded to self-efficacy items using a 1 (“Strongly Disagree”) to 7 (“Strongly Agree”) Likert scale, where higher scores indicate higher self-efficacy. Students completed items measuring self-efficacy once, after reviewing an NW9 example item but before engaging in subtest 1 of NW9 (see Appendix A for example item and proctor instructions when presenting this item).
Unfortunately, the reliability of the scores from the three self-efficacy items was less than adequate ($\alpha = .47^1$). Thus, the scores are not trustworthy. This finding resulted in removing self-efficacy from all proposed models (as explained below).

Data Analysis

Model-Data fit. To address the purpose of the present study, multiple models involving self-efficacy, perceived value, test emotions, motivation, and test performance were proposed. To test the global fit of each model, the $\chi^2$ goodness of fit (GOF) test was examined. In structural equation modeling, the $\chi^2$ GOF test is a standard method to test the fit of the model to the data (Hoyle & Panter, 1995); however, several limitations of the GOF test are known. First, the $\chi^2$ GOF test evaluates perfect model-data fit, meaning the hypothesis being tested is the model-implied data align perfectly with the sample data. Hoyle and Panter (1995) acknowledge that testing for perfectly-fitting models in social sciences is rarely plausible; thus, the test of perfect fit may be too restrictive. Second, as with any statistical test, the $\chi^2$ GOF test becomes over-powered with large samples resulting in minor model misfit being deemed statistically significant.

Thus, in addition to the $\chi^2$ GOF test, three approximate fit indices were evaluated based on simulation studies (Hu & Bentler, 1998, 1999): the Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Residual (SRMR). The CFI quantifies the difference in model fit between the estimated model and independence model. An independence model specifies that the variables in the model are not related; only the variances of each variable are estimated.

---

1 The response scale for self-efficacy items (1 to 7) was different from the response scale for all other items (1 to 5). Thus, it is possible that the differences in response scales have resulted in low reliability of self-efficacy scores.
Low CFI values may be obtained if the relationships between modeled variables are weak, and thus the independence model fits the data sufficiently well. The RMSEA is an absolute fit index, which quantifies the degree of model misspecification per degree of freedom. The SRMR is an absolute fit index that reflects the average correlation residual. A correlation residual is the difference between the observed correlation among two variables and the model-implied correlation. CFI values > .95, RMSEA values < .05, and SRMR values < .08 are suggested as indicative of adequate model-data fit (Hu & Bentler, 1999). However, these values should serve only as crude guidelines when evaluating model-data fit, not “Golden Rules” (Marsh, Hau, & Wen, 2004).

To supplement evaluation of global model-data fit, correlation residuals were examined. Correlation residuals expose local model-data misfit by indicating which relationships are not reproduced well and which relationships are reproduced well by the model. For example, if the observed relationship between effort and importance equals .5, and the model reproduces this relationship perfectly (i.e., .5), then the correlation residual would equal 0, indicating perfect model-data fit with respect to this relationship. Positive correlation residuals indicate that the model-implied correlation is an underestimate of the observed correlation, whereas negative correlation residuals indicate that the model-implied correlation is an overestimate of the observed correlation. Correlation residuals are an effect size on the metric of a correlation. There is no “cutoff” for large residuals, although some researchers indicate residuals greater than |.10| may be considered indicative of local misfit (Kline, 2011). For the current study, correlation residuals greater than |.10| will be noted and correlation residuals greater than |.15| will be deemed indicative of model-data misfit.
In addition to evaluating global and local model-data fit, a series of nested model comparisons were performed. That is, when models are nested models, the relative fit of the two models can be statistically tested using the $\chi^2$ GOF difference test. The degrees of freedom for the $\chi^2$ difference test is calculated by subtracting the degrees of freedom of a more complex model from the degrees of freedom of a simpler model. A simpler model (i.e., with more degrees of freedom) can never fit the data better than a more complex model (i.e., with less degrees of freedom); however, the two models can fit the data approximately the same (i.e., the difference in global fit is not statistically significantly different). The $\chi^2$ difference test suffers from the same limitations as the $\chi^2$ GOF test; thus, the change in fit indices was also evaluated (Cheung & Rensvold, 2002). Researchers recommend that CFI change of $.01$ (Cheung & Rensvold, 2002) and RMSEA change of $.01$ (Chen, 2007) may be indicative of significant change in model fit. Lastly, when comparing models, correlation residuals were examined. An increased number of large (i.e., greater than $.10$ or $.15$) residuals when moving from a more complex model to a simpler model suggests local fit becomes worse.

Autoregressive panel models

A series of panel models were proposed to test a series of hypotheses aligned with EV and CV theory as applied to test-taking motivation (see Table 7 for hypotheses and associated figures/models). The proposed models are under the subheading “Proposed autoregressive panel models”. Unfortunately, the proposed autoregressive panel models and associated research questions (Table 7) required modification since self-efficacy and perceived value at time 0 (prior to starting the test) had low reliability and could not be modeled. The actual estimated models are detailed under the subheading “Estimated
autoregressive panel models”. For the purposes of the thesis, it was important to include the proposed models even though these models could not be estimated. With that said, one can skip the proposed model text, if desired, and focus on the estimated model text. The proposed models that could not be estimated highlight the need for future study, which is discussed in the final chapter.

**Proposed autoregressive panel models.** First, two autoregressive models were proposed (Figure 6 and Figure 7). Autoregressive models (A and B) represent the hypothesis that there are no cross-lagged relationships; the constructs measured at different time points are only related via autoregressive effects. As specified in autoregressive model A, each construct at time t is allowed to relate to the same construct at time t+1. Thus, if this model fit the data, scores on a construct measured at a previous time (e.g., anger at time 1) are related to scores on a construct measured at a later time point (e.g., anger at time 2). However, constructs are not related within-time (e.g., anger 1, effort 1, importance 1) or across-time (e.g., anger at time 1, effort at time 2). If autoregressive model A fit the data, there would be no need to model any relations between value, effort, importance and performance, as these relations would be zero within and across-time.

As specified in autoregressive model B, correlations between disturbance terms of endogenous variables within a time point were estimated. These disturbance correlations represent that constructs are related to each other within-time, but are not related across-time. Thus, the only, yet important, difference between the autoregressive model A and autoregressive model B is the addition of 84 correlations between disturbances of constructs at time 1, 2, and 3 in the autoregressive model B. If autoregressive B fit the
data well, it would mean that each construct is related to the same construct at the previous time via the autoregressive effect (e.g., importance 1 to importance 2) and all constructs are related to each other within the same time. There would be no need to model any relations between different constructs across-time (e.g., direct effect from importance 1 to effort 2).

Both autoregressive model A and B can be considered baseline models because direct effects from previous time points are only estimated for the same construct (e.g., anger at time 1 directly affects anger at time 2). All cross-lagged effects (e.g., effect of anger at time 1 on effort at time 2) are constrained to equal 0. Thus, the two autoregressive models are the most parsimonious models tested.

Conducting a nested model comparison between autoregressive model A and autoregressive model B evaluates the importance of disturbance correlations. If autoregressive model B fits the data better than the autoregressive model A then each construct not only relates to the previous measure of that construct via autoregressive effects, but all constructs at a particular timepoint relate to each other. Finding these disturbance correlations to be significant is critical as it indicates that constructs within the same time are systematically affected by other constructs, and thus by not including these effects, the estimated model is misspecified.

I hypothesize that autoregressive model B will fit the data better than the autoregressive model A because theoretically the constructs are hypothesized to relate to each other, and the within-time relations will be represented by the disturbance correlations. The absolute fit of both models will not be adequate because, in addition to the within-time relations, the constructs are also hypothesized to have across-time (e.g.,
cross-lagged) relations. Autoregressive model B is used as the baseline model for further nested model comparisons discussed below.

Next, two EV theory models (A and B) were proposed (Figure 8 and Figure 9). These models are considered EV models because the cross-lagged effects of perceived value and self-efficacy on examinee effort in these models align with EV theory. That is, as specified in EV model A, the cross-lagged effect of perceived value at time \( t \) on examinee effort at time \( t+1 \) are estimated. Similarly, the effect of self-efficacy at time 0 on examinee effort at time 1 are estimated. Both of these cross-lagged effects represent that the antecedents of examinee effort are perceived value and self-efficacy. As specified in EV model B, the effects of self-efficacy at time 0 on examinee effort at time 2 and time 3 are estimated. According to EV theory, self-efficacy and perceived value should affect subsequent examinee effort. However, in this study, additional measurements of self-efficacy were not feasible; thus, when modeling examinee effort at time 2 and time 3, the effects of self-efficacy measured at time 0 were estimated. Note that using this rationale, many of the following models (termed “B”) contain estimates of these effects. Lastly, the effects of examinee effort at time \( t \) on test performance at time \( t+1 \) were estimated.

Comparing EV theory model A to autoregressive model B tests the hypothesis that the effects of situational appraisals (i.e., value and self-efficacy) on subsequent motivation, and the effects of motivation on subsequent performance, are important. I hypothesize that EV theory model A will fit the data better than autoregressive model B. Next, I will compare the EV theory model A and EV theory model B. This comparison tests if the effects of self-efficacy at time 0 on motivation at time 2 and time 3 are important, after controlling for motivation at time 1 and time 2 respectively. If EV theory model B fits
better than the EV theory model A then EV theory model B will serve as baseline for next model comparisons. Otherwise, EV theory model A will be used as baseline model. Finding that both EV theory models fit the data well would garner support for EV theory. However, I hypothesize that both of these models will fit the data poorly in an absolute sense, because other constructs (e.g., test emotions) are hypothesized to have effects on examinee effort and test performance.

Next, two Combined EV and CV theory models (A and B) were proposed (Figure 10 and 11). The Combined EV and CV theory model A builds on the EV theory model A by estimating the effects of perceived value and self-efficacy at time t on anger, joy, pride, worry, and boredom at time \( t+1 \) and the effects of test emotions at time t on motivation at time \( t+1 \). The Combined EV and CV theory model B builds on the EV theory model B the same way – by estimating the effects of perceived value and self-efficacy at time t on anger, joy, pride, worry, and boredom at time \( t+1 \) and the effects of test emotions at time t on motivation at time \( t+1 \). These two models specify both CV and EV theory effects.

To evaluate the importance of the CV theory effects, two model comparisons were proposed. First, the Combined EV and CV theory model A should be compared to the EV theory model A. Second, the Combined EV and CV theory model B should be compared to the EV theory model B. By comparing the EV theory models (which constrain the effects of value and self-efficacy on subsequent emotions) to the Combined EV and CV theory models (which estimate the effects of value and self-efficacy on subsequent emotions), the CV theory effects of perceived value and self-efficacy on subsequent emotions could be evaluated in the presence of EV theory effects. If EV
theory is correct (i.e., examinee effort is predicted by perceived value and self-efficacy), then the CV theory effects will be negligible and the Combined EV and CV theory models will fit the data as well as the EV theory models. Otherwise, CV theory effects are important to consider.

I hypothesize that the Combined EV and CV theory models (A and B) will fit data better than the EV theory models (A and B), respectively. There are two ways in which the Combined EV and CV theory models are different from the EV theory models. First, the Combined EV and CV theory models estimate the effects of appraisals on subsequent emotions. Second, Combined EV and CV theory models estimate the effects of emotions on subsequent motivation. Thus, two follow-up models were estimated to diagnose which of these effects are important. The first model estimated the effects of appraisals on subsequent emotions. The second model estimated the effects of emotions on subsequent motivation. These models are not nested with each other; however, they are nested within the more complex Combined EV and CV theory model. By comparing these two models to the Combined EV and CV theory model individually, the importance of appraisals on subsequent emotions, and the importance of emotions on subsequent motivation will be evaluated. If Combined EV and CV theory model B fits the data better than the Combined EV and CV theory model A, then the two follow-up models will be built on Combined EV and CV theory model B. Given I hypothesize that the Combined EV and CV theory models will fit the data better than the EV theory, the Combined EV and CV theory models will serve as baseline models for the next set of nested model comparisons.
Next, CV theory models (A and B; Figure 12 and 13) were proposed. In CV theory model A, the effects of value time t on test emotions at time t+1, and the effects of self-efficacy at time 0 on test emotions at time 1 were estimated. Additionally, the effects of test emotions at time t on motivation at time t+1, and the effects of motivation at time t on performance at time t+1 were estimated. In CV theory model B, the effects of self-efficacy at time 0 on motivation at time 2 and time 3 were estimated. The comparison between CV theory model A and CV theory model B provided a more rigorous test for the self-efficacy effects on subsequent motivation.

By comparing the CV theory models (which constrain the effects of value and self-efficacy on subsequent examinee effort) to the Combined EV and CV theory models (which estimate the effects of value and self-efficacy on subsequent examinee effort), the EV theory effects of perceived value and self-efficacy on subsequent examinee effort could be evaluated, in the presence of CV theory effects. If CV theory is correct (i.e., examinee effort is predicted by test emotions), then the EV theory effects on effort will be negligible and the Combined EV and CV theory models will fit the data as well as the CV theory models. Otherwise, EV theory effects are important to consider. The comparison of EV, CV, and the Combined EV and CV theory models address the main purpose of the study – examining the antecedents of examinee effort. As outlined above, by comparing these models, the effects of EV theory and CV theory were evaluated in the presence of the effects from both theories, resulting in a rigorous test for each theory.

It was expected that the six models described above (EV theory model A and B, CV theory model A and B, Combined EV and CV theory model A and B) would result in different model-data fit. However, in terms of absolute model data fit, it was expected
that none of six models would fit the data adequately because reciprocal effects were fixed to zero. According to CV theory, reciprocal effects between perceived value, self-efficacy, test emotions, examinee effort, and test performance exist. Therefore, an additional four models were proposed. Given I hypothesize that CV theory models will fit as well as the Combined EV and CV theory models, the two CV theory models will be used as baseline models for next set of model comparisons.

I specified two Adjacent Reciprocal Effects Models based on CV theory (Figure 14 and 15). According to CV theory, situational appraisals (perceived value and self-efficacy) are hypothesized to affect subsequent emotions, which then affect subsequent examinee effort, which then affect subsequent test performance. The Adjacent Reciprocal Effects Models specify the reciprocal effects of the adjacent variables in this CV causal chain. Adjacent Reciprocal Effects Model A tests the effects of emotions on subsequent perceived value, the effects of examinee effort on subsequent emotions, and the effect test performance on subsequent examinee effort. If this model fits the data better than CV model A, then such finding would suggest adjacent reciprocal effects are important to consider.

Second, the Adjacent Reciprocal Effects Model B specifies the effects of self-efficacy at time 0 on subsequent emotions at time 2 and time 3. The comparison between Adjacent Reciprocal Effects Model A and the Adjacent Reciprocal Effects Model B, tests the importance of the effects of self-efficacy at time 0 on test emotions at time 2 and time 3. I hypothesize that Adjacent Reciprocal Effects Models will fit the data better than the CV theory model. However, neither model will fit the data well in an absolute sense.
because according to CV theory, reciprocal effects between all constructs in the CV causal chain exist.

Last, Complete Reciprocal Effects Models (A and B, Figures 16 and 17) were proposed. Building on the Adjacent Reciprocal Effects Model A, the Complete Reciprocal Effects Model A specifies the effects of situational appraisals on subsequent motivation and test performance, the effects of test emotions on subsequent performance, the effects of motivation on subsequent value, the effects of test performance on subsequent value and subsequent test emotions. Complete Reciprocal Effects Model B builds on the Complete Reciprocal Effects Model A by specifying the effects of self-efficacy at time 0 on test emotions at time 2 and time 3. The comparison between Complete Reciprocal Effects Model A and Complete Reciprocal Effects Model B tests if the effects of self-efficacy at time 0 on test emotions at time 2 and time 3 are important.

I hypothesize that both of these models will fit well in an absolute sense, since they align with CV theory, which specifies the reciprocal effects between all constructs in the CV causal chain. Comparing the Complete Reciprocal Effects Models to the Adjacent Reciprocal Effects Models will provide a test for the reciprocal effects that are more distal within the CV theory causal chain (i.e., situational appraisals impact test emotions, which then impacts motivation, which then influences test performance).

**Estimated autoregressive panel models.** The proposed autoregressive panel models and associated research questions (Table 7) detailed above were modified since self-efficacy and perceived value at time 0 were no longer modeled due to low reliability. Given this, the estimated autoregressive panel models and the associated research
questions changed and are summarized in Table 8. The new, modified autoregressive panel models are discussed below.

First, two autoregressive models (A and B) were estimated (Figure 18 and Figure 19). As specified in Autoregressive Model A, the constructs at time t were only allowed to relate to the same construct at time t+1. Thus, Autoregressive Model A represents a hypothesis that a construct only relates to the same construct at a subsequent time—all other correlations were expected to be zero. That is, no across-time relations between constructs (e.g., anger at time 1 and pride at time 2) were hypothesized to exist. Constructs were allowed to relate within-time only at time 1, as these were exogenous variables that began the longitudinal panel model. Within-time relations between constructs at time 2 and time 3 were set to zero. If Model A fit the data well, then no further models would need to be estimated.

As specified in Autoregressive Model B, correlations between disturbance terms of constructs at time 2 and time 3 were estimated. Autoregressive Model B represents a hypothesis that constructs are related within-time, which is represented by the 56 disturbance correlations at time 1 and 2 and 28 correlations at time 1. If disturbance correlations were significant, it would indicate systematic variance between within-time constructs.

By comparing the two autoregressive models, the importance of disturbance correlations (i.e., the importance of within-time relations) was evaluated. I hypothesized that the Autoregressive Model B will fit the data better than the Autoregressive Model A because constructs are theorized to be related within the same time point. However, neither of the models will fit the data well in an absolute sense because, in addition to the
within-time relations, constructs are also hypothesized to have across-time (i.e., cross-lagged) relations. Autoregressive Model B was used as the baseline model for further model comparison.

Next, the EV Theory Model was estimated (Figure 20). This model built on Autoregressive Model B by estimating two cross-lagged effects: the effect of perceived value at time t on examinee effort at time t +1 and the effect of examinee effort at time t on test performance at time t +1. Both of these effects align with EV theory in that perceived value of a test should influence subsequent motivation to complete a test and motivation should influence subsequent performance. However, the EV Theory Model did not completely represent EV theory because the effects of self-efficacy (and the effects of the interaction between self-efficacy and perceived value) on subsequent motivation were not included due to poor measurement. Thus, finding that the EV Theory Model fit the data would lend only partial support for EV theory. I hypothesized that the EV Theory Model would not fit the data well in an absolute sense because other constructs (e.g., test emotions) were hypothesized to have cross-lagged effects on examinee effort. By comparing the EV Theory Model and the Autoregressive Model B, the EV theory effects (i.e., the effects of perceived value on subsequent effort and the effects of effort on subsequent performance) were evaluated. I hypothesized that these cross-lagged effects are important and thus, EV Theory Model would fit the data better than Autoregressive Model B, which constrained these effects to equal zero. The EV Theory Model was used as the baseline model for next model comparisons.

Next, the Combined EV and CV Theory Model was estimated (Figure 21). The Combined EV and CV Theory Model was built on the EV Theory Model by estimating
the effects of perceived value at time $t$ on anger, joy, pride, worry, and boredom at time $t+1$ and the effects of test emotions at time $t$ on motivation at time $t+1$. The Combined EV and CV theory Model evaluated the effects based on both EV and CV theories. Finding that this model fit the data well would suggest that students’ motivation is influenced both by perceived value and by test emotions, while controlling for each other. The Combined EV and CV Theory Model should fit the data better than the EV Theory Model given two key differences between the models. First, the Combined EV and CV Theory Model estimated the effects of perceived value on subsequent emotions. Second, the Combined EV and CV Theory Model estimated the effects of emotions on subsequent motivation. These additional effects align with CV theory. Thus, by comparing the Combined EV and CV theory Model and the EV Theory Model, the importance of CV theory effects in the presence of EV theory effects was evaluated.

Next, the CV Theory Model was estimated and compared to the Combined EV and CV Theory Model to examine the importance of the EV effects (Figure 22). The importance of EV theory effects were already evaluated by comparing the EV Theory Model and the Autoregressive Model B. However, by comparing the Combined EV and CV Theory Model and the CV Theory Model, the EV theory effects are evaluated in the presence of the CV theory effects, providing a more rigorous test of the EV theory effects. The CV Theory Model was created by setting the Combined EV and CV Theory Model effects of perceived value at time $t$ on examinee effort at time $t+1$ to zero. The CV Theory Model estimated the effects of perceived value at time $t$ on test emotions at time $t+1$, the effects of test emotions at time $t$ on motivation at time $t+1$, and the effects of motivation at time $t$ on performance at time $t+1$. 
One of the purposes of this study was to compare two theories of students’ motivation: EV theory and CV theory. The EV Theory Model and the CV Theory Model are not nested within each other; thus, a nested model comparison between these two models could not be performed. However, both of these models were evaluated with respect to a more complex Combined EV and CV Theory Model. Comparing the EV Theory Model and the Combined EV and CV Theory Model evaluated the importance of CV theory effects, in the presence of EV theory effects. Comparing the CV Theory Model and the Combined EV and CV Theory Model evaluated the importance of EV theory effects, in the presence of CV theory effects. I hypothesized that the CV Theory Model would not fit the data worse than the Combined EV and CV Theory Model, indicating the EV effects of perceived value on subsequent examinee effort could be fixed to zero. However, the EV Theory Model would fit the data worse than the Combined EV and CV Theory Model, indicating the CV effects of perceived value on subsequent test emotions and the effects of test emotions on subsequent examinee effort are not nil. Given that the CV Theory Model was more parsimonious than the Combined EV and CV Theory Model, the CV theory model would then be used for further model comparisons.

None of the three models listed above (EV Theory Model, CV Theory Model, or the Combined EV and CV Theory Model) were hypothesized to fit the data well in an absolute sense. The EV Theory Model did not estimate the effects of perceived value on subsequent test emotions nor the effects of test emotions on subsequent examinee effort (i.e., CV theory effects). Thus, in comparison to the CV Theory Model and the Combined EV and CV Theory Model, the EV Theory Model was misspecified. The CV Theory
Model and the Combined EV and CV Theory Model estimated the non-negligible CV effects; however, both were hypothesized to be misspecified. According to CV theory, perceived value, test emotions, motivation, and test performance are related via reciprocal effects. Thus, two competing models that include reciprocal effects were estimated.

Building on the CV Theory Model, the Adjacent Reciprocal Effects Model (see Figure 23) estimated the effects of test emotions at time t on perceived value at time t+1, the effects of motivation at time t on test emotions at time t+1, and the effects of test performance at time t on motivation at time t+1. These effects were estimated because they represent the reciprocal effects between adjacent variables in the CV causal chain (Figure 2). I hypothesized that the Adjacent Reciprocal Effects Model would fit the data better than the CV theory model, suggesting that reciprocal effects are important to consider. However, the absolute fit of the Adjacent Reciprocal Effects Model should not be adequate because according to CV theory, reciprocal effects between all constructs in the CV theory causal chain should be estimated.

Thus, the Complete Reciprocal Effects Model was also estimated (Figure 24). Building on the Adjacent Reciprocal Effects Model, the effects of perceived value at time t on motivation and test performance at time t+1, the effects of test emotions at time t on test performance at time t+1, the effects of motivation at time t on perceived value at time t+1, and the effects of test performance at time t on perceived value and test emotions at time t+1, were estimated. Comparing the fit of the Complete Reciprocal Effects Model and the Adjacent Reciprocal Effects Model provided a test of the reciprocal effects that are more distal within the CV theory causal chain. I hypothesized that the Complete Reciprocal Effects Model would fit the data better than the Adjacent Reciprocal Effects
Model because according to CV theory all constructs are related via reciprocal effects. I also hypothesized that the Complete Reciprocal Effects Model would fit the data well in an absolute sense since the Complete Reciprocal Effects Model aligns with CV theory. Thus, the Complete Reciprocal Effects Model should fit the data the best out of all the models tested in the current study.
Chapter 4: Results

Before estimating the autoregressive panel models, the data were examined descriptively and both stability and equilibrium were assessed. Then the panel models were compared. The results from these comparisons were structured according to the research questions.

Descriptive Statistics

Descriptive statistics were obtained (mean, standard deviation, skewness, kurtosis, and coefficient α) for all 26 variables in the current study (Table 6). The possible range of perceived value and examinee effort is 5 to 25, which was observed at all but one measurement occasion (at time 1, the observed range of examinee effort was 6 to 25). The possible range of all test emotions is 3 to 15, which was observed at all measurement occasions. The possible range for each subtest is 0 to 22, whereas the observed range of scores was 5 to 22 for subtest 1, 3 to 22 for subtest 2, and 3 to 21 for subtest 3. Across time, the means of perceived value and examinee effort ranged from 16.25 and 18.09, indicating moderate levels of each construct (i.e., if scaled from 1 to 5, the means range from 3.25 to 3.62). Across time, the means of each emotions ranged from 6.41 to 10.19, indicating low to moderate levels of each emotions (i.e., if scaled from 1 to 5, the means range from 2.14 to 3.40). Across time, the means for each subtest ranged from 14.04 to 15.01, indicating students answered about 64% of items correctly in each subtest.

Next, univariate and multivariate normality were checked for all variables. Univariate skewness and kurtosis values greater than |2| and |7| may be indicative of non-normality (Finney, DiStefano, & Kopp, 2016). No values reached these cut-offs, suggesting that univariate normality was not violated. Multivariate normality was
evaluated using Mardia’s normalized multivariate kurtosis statistic. No clear cut-off value has been established (Finney & DiStefano, 2013). Mardia’s standardized kurtosis was 79.78 in the current sample.

The autoregressive models were estimated using both Maximum Likelihood (ML) estimation and Maximum Likelihood estimation with Satorra-Bentler adjustments (Satorra & Bentler, 1994). The conclusions did not differ across estimators; thus, ML results are reported. I attempted to estimate the autoregressive models using a single-indicator latent variable approach (Cole & Preacher, 2014) as well; however, many models did not converge to an admissible solution. Path model results, with no correction for unreliability, are presented and interpreted.

**Stability**

Before estimating the proposed autoregressive panel models, stability of the constructs was examined. Recall, stability is the opposite of change (Cole & Maxwell, 2003). Stability can refer to the equality of mean scores of a construct over time (i.e., absolute stability) or high autocorrelations between the same construct over time (i.e., relative stability) (Newsom, 2015).

First, absolute stability was evaluated by estimating a model that constrained the mean of a construct at time 1 to equal the mean of the construct at time 2 and time 3. For example, the mean level of perceived value at time 1 was constrained to equal the mean level of perceived value at time 2 and the mean level of perceived value at time 3. An “omnibus” model constraining the means of all 8 constructs to be equal over time was estimated, resulting in 16 degrees of freedom. This model did not fit the data well, indicating the means were not equal over time: $\chi^2 (16) = 464.422$, $p < .0001$, RMSEA =
.171, CFI = .979. Next, a series of models constraining the means of each construct to be equal were estimated. The results for each model are reported in Table 9. Only the mean level of boredom was stable throughout the test. The mean levels of the other seven constructs changed throughout the test. The means of perceived value, effort, worry, and pride decreased over time, whereas the means of boredom and anger increased over time. The mean of enjoyment decreased from time 1 to time 2, but then slightly increased from time 2 to time 3. The mean of test performance increased from time 1 to time 2 but then decreased from time 2 to time 3 (Table 6).

Second, relative stability was evaluated. Relative stability refers to the unchanging rank order of individual scores over time (Cole & Maxwell, 2003; Newsom, 2015). If the correlation between a construct at time 1 and time 2 is equal to 1, then the rank order of individuals is not changing over time; individual differences on the construct are completely stable over time. It is impossible for another construct to have an effect on a stable construct. In other words, after controlling for the same construct at a previous time, there are no individual differences left for other constructs to predict. One approach to take when examining relative stability is to evaluate the bivariate correlations between the same construct over time. Although none of the bivariate correlations equal 1 (see Table 10), some values are approaching 1 (e.g., correlation between Worry at time 2 and Worry at time 3 is .882).

A similar approach to testing relative stability is to use regression analysis to predict the construct at time 2 from the construct at time 1 and assess if the error variance is near zero (Newsom, 2015). As specified in the Autoregressive Model A, the effects of

\[ \text{No structure among the constructs (i.e., no effects or relationships between constructs) was imposed; thus, calculating SRMR is not necessary.} \]
a construct at time t on the construct at time t+1 are estimated, resulting in 16 autoregressive effects and 16 error variances. Estimating this model using LISREL 9.3 (Scientific Software International) provides a statistical test for each error variance from a value of 0. All of the error variances were significantly greater than zero, providing further evidence that the constructs in the study do not display perfect relative stability. In sum, neither absolute stability (i.e., mean levels of a construct) nor relative stability (i.e., high autocorrelations) were observed, indicating constructs were changing throughout the test and had the potential to be predicted by other constructs. However, since the majority of autocorrelations between constructs were high (i.e., ranging from .70 to .88), the cross-lagged effects of a construct on a different construct may be small in magnitude. That is, the higher the autocorrelation between the same construct over time, the less variance there is remaining to be explained by the predictors. Thus, if the unstandardized and standardized cross-lagged effects of one construct on another construct are small, this may be due to the construct not changing much over time (hence the large autoregressive effect). However, even small cross-lagged effects may be statistically and practically significant.

**Equilibrium**

“The latent variable system is in equilibrium if the variances and covariances of the latent variables are invariant from one wave to the next” (Cole & Maxwell, 2003, p. 570). Establishing equilibrium means any changes in the causal system have already manifested their effects and the system is in a steady state (Kline, 2011). The causal process has dampened out; it is not just beginning (Kenny, 1979). The lack of equilibrium
implies that constructs not in the model may be affecting each construct differently at different times, which may warrant further attention.

Equilibrium is assessed by testing equivalent cross-sectional variances and covariances across time (Cole & Maxwell, 2003; Preacher, 2015). In the current study, only the equality of cross-sectional variances was tested by estimating a model constraining the variances of each construct to be equal across time. That is, the variance of perceived value at time 1 was constrained to equal the variance of perceived value at time 2 and time 3. An “omnibus” test constraining the variances of all 8 constructs was then fit to the data. This model did not fit the data well: $\chi^2 (16) = 173.499, p < .0001$, RMSEA = .108, CFI = .993, SRMR = .0445. Next, eight models were estimated where the variances of each construct were constrained to be equal (see Table 11). The results suggest that only the variance of boredom is equal across the three measurement occasions. The variances of perceived value, effort, pride, anger, and enjoyment increased over time, whereas the variance for worry decreased over time (Table 6). Lastly, the variance of test performance increased from time 1 to time 2 but decreased from time 2 to time 3.

The changes in variance across time are important to consider when interpreting the unstandardized paths and disturbance terms in the panel models. If the unstandardized effects of one construct on another construct are equivalent across time, but the variance of these two constructs change over time, then the amount of error variance associated with each endogenous variable (i.e., disturbance term) will change over time. Similarly, the standardized effects will change. Thus, unequal variances will become important to consider when discussing stationarity of direct effects below. In sum, equilibrium was not
supported, which in not surprising. Equilibrium is rarely achieved in social sciences; thus, these results should not be discouraging (Jose, 2016).

**Autoregressive panel models**

**Research question 1.** Can value, effort, emotion, and performance relations across time be adequately explained by autoregressive effects? Autoregressive Model A represents a hypothesis that after controlling for the autoregressive effects associated with each construct, there are no within-time (except at time 1) or across-time (i.e., cross-lagged) relations between value, effort, test emotions, and test performance (see Table 12). If Autoregressive Model A fit the data well in an absolute sense, then no further models would need to be estimated or tested.

This model did not fit the data well: $\chi^2 (232) = 2186.895$, $p < .0001$, CFI = .907, RMSEA = .094, SRMR = .116. When examining areas of local misfit, 39 relationships were reproduced poorly (i.e., correlation residuals > |.15|), and other 74 relationships were not reproduced well (i.e., correlation residuals were > |.10|). Of the 28 within-time relationships between constructs at time 2, 23 correlation residuals were greater than |.10|. Of the 28 within-time relationships between constructs at time 3, 25 correlation residuals were greater than |.10|.

**Research question 2.** Can value, effort, emotion, and performance relations across time be adequately explained via autoregressive effects and *within-time* correlations? Given the poor fit of Autoregressive Model A due to setting the within-time correlations to zero, Autoregressive Model B freely estimated these within-time relations. Thus, the two autoregressive models were identical except for 56 within-time disturbance correlations at time 2 and time 3. Statistically comparing the two models formally tests
the importance of within-time bivariate relations. As expected, Autoregressive Model B fit the data better than the Autoregressive Model A: \( \Delta \chi^2 (56) = 1035.727, p < .0001, \Delta \text{CFI} = .047, \Delta \text{RMSEA} = .018. \)

Although Autoregressive Model B did not fit the data well in an absolute sense (see Table 12), examining the correlated disturbances provides insight into the shared variability between the within-time variables after controlling for the autoregressive effects. Of the 56 disturbance correlations, 37 were statistically significantly different from zero, suggesting the need to account for the within-time variability either via disturbance correlations (as done in Autoregressive Model B) or via direct effects of other variables (as done in the subsequent panel models). Thus, the change in disturbance correlations when testing the panels models will uncover if these within-time relations are explained by emotions, effort, value and performance experienced earlier during the test.

When examining local model misfit, both autoregressive models produced approximately the same number of problematic correlation residuals (i.e., > |.15|). As expected, the number of within-time correlation residuals greater than |.10| decreased for the Autoregressive Model B suggesting that the fit improved for Autoregressive Model B. That is, of the 28 within-time relationships between constructs at time 2, 15 correlation residuals were greater than |.10|. Of the 28 within-time relationships between constructs at time 3, 22 correlation residuals were greater than |.10|. The number of across-time correlation residuals remained approximately the same when moving from Autoregressive Model A to Autoregressive Model B, suggesting the need to estimate across-time effects to reproduce across-time relations, as specified in the panel models tested next.
**Research question 3.** Can EV theory explain relations between value, effort, emotions, and performance across time? The EV Theory Model built on Autoregressive Model B by estimating the effect of perceived value on subsequent effort and the effect of effort on subsequent test performance. The EV Theory Model fit the data significantly better than Autoregressive Model B: $\Delta \chi^2 (4) = 83.761, p < .0001$. However, the change in approximate fit suggested the models were not practically different ($\Delta$CFI = .003, $\Delta$RMSEA = .002). The SRMR was lower for the EV Theory Model, suggesting that, on average, the EV Theory Model reproduced the relationships between constructs better than Autoregressive Model B. When examining local model misfit, the number of problematic correlation residuals (i.e., $>|.15|$) decreased when moving from Autoregressive Model B (24) to the EV Theory Model (13). Thus, I deemed the EV Model as better fitting than Autoregressive Model B. The EV Model was retained for subsequent nested model comparisons.

However, neither model fit the data well in an absolute sense (Table 12). Recall, the EV Theory Model did not estimate direct paths between emotions and perceived value or direct paths between emotions and effort. Lastly, no direct paths between different emotions were specified. Thus, the EV Theory Model did not reproduce many of the relations between these constructs well. For example, the correlation between pride at time 3 and perceived value at time 3 was underestimated by .16, whereas the correlation between pride at time 3 and examinee effort at time 3 were underestimated by .23. Additionally, relations between emotions within and across time were estimated poorly. For example, the correlation between worry at time 3 and anger at time 3 was
underestimated by .17, whereas the relationship between anger at time 2 and worry at time 3 was underestimated by .17.

**Research question 4.** Can the Combined EV and CV Theory Model explain the relations between value, effort, emotions, and performance across time? The Combined EV and CV Theory Model built on the EV Theory Model by estimating the effects of perceived value on subsequent test emotions and the effects of test emotions on subsequent examinee effort. Thus, by comparing the Combined EV and CV Theory Model to the EV Theory Model, the importance of the CV theory effects was tested (see Table 12). The Combined EV and CV theory model fit the data significantly better than the EV theory Model: $\Delta \chi^2 (20) = 154.783, p < .0001$. Although the change in CFI and RMSEA did not suggest that the two models are practically different ($\Delta$CFI = .007, $\Delta$RMSEA = .001), the SRMR was noticeably lower for the Combined EV and CV Theory Model (.069) than for the EV Theory Model (.101). Related, the Combined EV and CV Theory Model reproduced relations between constructs better than the EV Theory Model. The Combined EV and CV Theory Model produced only three large correlation residuals (i.e., $>|.15|$), whereas the EV Theory Model produced 13 large correlation residuals.

Summarizing the correlation residuals that decreased when moving from the EV Theory Model to the Combined EV and CV Theory Model helps to inform discussion of the importance of the CV effects. That is, recall the difference between the EV Theory Model and the Combined EV and CV Theory Model is the estimation of the effects of perceived value on subsequent test emotions and the effects of test emotions on subsequent examinee effort. Thus, it was of interest to see if the correlation residuals
between these relations decreased when comparing the EV Theory Model to the Combined EV and CV Theory Model. When comparing local misfit from both models, three correlation residuals decreased from being greater than |.15| to being lower than |.10|: pride at time 2 and effort at time 2, pride at time 2 and effort at time 3, perceived value at time 3 and pride at time 3. Seven correlation residuals decreased from being greater than |.15| to being greater than |.10|. Of these seven correlation residuals, the biggest decreases were observed for the relation between pride at time 3 and effort at time 3 (residual of .23 reduced to .10) and the relation between effort at time 2 and pride at time 3 (residual of .21 reduce to .12). The decreases in these correlation residuals illustrate that the CV theory effects (estimated in the Combined EV and CV Theory Model) and non-negligible and improve local model-data fit. The remaining five correlation residuals were between emotions within-time and across time. For example, the correlation residual between anger at time 3 and boredom at time 3 decreased from .17 to .12 and the correlation residual between anger at time 2 and boredom at time 3 decreased from .15 to .11. Thus, the change in correlation residuals from the EV Theory Model to Combined EV and CV Theory Model highlight the importance of CV theory effects.

The three correlation residuals that were the greatest in magnitude remained approximately the same when moving from EV Theory Model to the Combined EV and CV Theory Model (the greatest change was .005). These correlation residuals were for the relations between test performance at time 1 and test performance at time 3 (.19), anger at time 3 and worry at time 3 (.18), and between anger at time 2 and worry at time 3 (.17). In both models, there were no direct effects estimated between these variables.
Thus, although the Combined EV and CV Theory Model did not fit the data well in an absolute sense ($\chi^2 (152) = 912.624$, CFI = .964, RMSEA = .072, SRMR = .069), it fit better than the EV theory Model, indicating that CV theory effects are not negligible and are important to consider when predicting examinee effort. The obvious next question concerns if the EV theory effects are important when modeling the CV theory effects, which I address next.

**Research question 5.** Are the EV theory effects important when predicting examinee effort? Comparing the Combined EV and CV Theory Model and the CV Theory Model tested the importance of the EV theory effects. The Combined EV and CV theory model fit the data significantly better than the CV Theory Model: $\Delta \chi^2 (2) = 16.68$, $p < .0002$. However, the change in fit indices did not suggest the models are practically different ($\Delta\text{CFI} = .001$, $\Delta\text{RMSEA} < .001$). Similarly, the SRMR values were nearly identical for the Combined EV and CV Theory Model (.069) and the EV Theory Model (.074). Lastly, the number of worrisome correlation residuals (i.e., either $>|.15|$ or $>|.10|$) were similar across the two models.

The Combined EV and CV Theory Model resulted in three correlation residuals that were non-negligible (i.e., $>|.15|$), whereas the CV theory model resulted four correlation residuals that were non-negligible. The three correlation residuals that were greater than $|.15|$ for the Combined EV and CV Theory Model were observed for the CV Theory Model as well. The fourth correlation residual that was identified as problematic by the CV Theory Model was between effort at time 2 and importance at time 3 (residual = .158). The Combined EV and CV Theory Model also underestimated this relation by
.120. Thus, both models poorly reproduced this relation. However, only the CV Theory Model produced a residual that exceeded the |.15| cutoff.

In sum, the Combined EV and CV Theory Model and CV Theory Model were statistically different from each other. However, according to the fit indices and the examination of local model misfit, it appears the two models fit the data equally well. Since the CV Theory Model is simpler (i.e., fewer parameters are estimated) than the Combined EV and CV Theory Model and the few remaining correlation residuals were the same across the two models, the CV Theory Model was retained for subsequent nested model comparisons. Regarding the research question, EV theory effects may not be important when predicting subsequent examinee effort.

**Research question 6.** Can the CV Theory Model explain the relations between value, effort, test emotions, and test performance across time? Given the more complex Combined Model fit the data the same as the CV Theory Model, the absolute model data fit of CV Theory Model was tested. The CV Theory Model did not fit the data well: $\chi^2 (154) = 929.304, p < .0001$, RMSEA = .073, CFI = .963, SRMR = .074. As summarized above, four correlation residuals were problematic (i.e., > |.15|) and 41 were greater than |.10|. Thus, to account for some of the local model misfit, reciprocal effects were added to the CV Theory Model, which served as the baseline model in the next series of nested model comparisons.

**Research question 7.** Can Reciprocal Effects Models explain the relations between value, effort, test emotions, and test performance across time? To answer this research question, three nested models were compared: the CV Theory Model (Figure 20), the Adjacent Reciprocal Effects Model (Figure 23), and the Complete Reciprocal
Effects Model (Figure 24). The Adjacent Reciprocal Effects Model built on the CV Theory Model by estimating effects between constructs that are “adjacent” to each other in the CV theory causal chain (Figure 2). Thus, by comparing the Adjacent Reciprocal Effects Model to the more parsimonious CV Theory Model, the importance of the following effects was evaluated: the effects of test emotions on subsequent perceived value, the effects of examinee effort on subsequent test emotions, and the effects of test performance on subsequent examinee effort.

The Adjacent Reciprocal Effects Model fit the data better than the CV Theory Model: \( \Delta \chi^2 (22) = 159.214, p < .0001 \). Although, the change in CFI and RMSEA indicated that the two models fit similarly (\( \Delta \text{CFI} = .007, \Delta \text{RMSEA} = .0015 \)), the SRMR value was noticeably lower for the Adjacent Reciprocal Effects Model (.049) than the CV Theory Model (.074). Likewise, the Adjacent Reciprocal Effects Model has fewer and smaller correlation residuals than the CV Theory Model. Only two correlation residuals were problematic (i.e., > |.15|) for the Adjacent Reciprocal Effects Model, whereas four correlation residuals were problematic for the CV Theory Model. Similarly, 13 relationships were not reproduced well (i.e., correlation residuals > |.1|) by the Adjacent Reciprocal Model, whereas 41 relationships were not reproduced well by the CV Theory Model. In sum, the Adjacent Reciprocal Effects Model fit the data better globally and locally than the CV Theory Model. Moreover, the Adjacent Reciprocal Model fit well in an absolute sense (Table 12).

The Adjacent Reciprocal Effects Model was compared to the Complete Reciprocal Effects Model. Building on Adjacent Reciprocal Effects Model, the Complete Reciprocal Effects Model estimated reciprocal effects between all constructs in the CV
theory causal chain (Figure 2). By comparing the Complete Reciprocal Effects Model to the Adjacent Reciprocal Effects Model, the importance of the remaining (i.e., not adjacent) reciprocal effects was evaluated: the effects of perceived value on subsequent examinee effort and test performance, the effects of test emotions on subsequent test performance, the effects of examinee effort on subsequent perceived value, and the effects of test performance on subsequent perceived value and test emotions.

As hypothesized, the Complete Reciprocal Effects Model fit the data significantly better than the Adjacent Reciprocal Effects Model: Δχ² (28) = 69.946, p < .0001. However, the change in the fit indices (i.e., RMSEA, CFI, and SRMR) was minimal between the two models, suggesting that the two models were equivalent with respect to model-data fit. Moreover, the pattern of local misfit was nearly identical for the two models. For the Adjacent Reciprocal Effects Model, two correlation residuals were problematic (i.e., > |.15|), whereas for the Complete Reciprocal Effects Model, one correlation residual was problematic. Similarly, for the Adjacent Reciprocal Effects Model, 13 relationships were not reproduced well (i.e., correlation residuals > |.10|), whereas for the Complete Reciprocal Effects Model, 15 relationships were not reproduced well.

Moreover, the pattern of local model misfit in both models was the same. That is, the same relationships between constructs were either over-or-under reproduced by both models. The highest correlation residual (in absolute value) in both models was between test performance at time 1 and test performance at time 3. The Adjacent Reciprocal Effects Model underestimated the relationship by .175, whereas the Complete Reciprocal Effects Model underestimated the relationship by .158. The second highest correlation
residual was between anger at time 3 and worry at time 3. The Adjacent Reciprocal Effects Model underestimated the relationship by .151, whereas the Complete Reciprocal Effects Model underestimated the relationship by .147. Although Complete Reciprocal Effects Model produced one correlation residual greater than .15 (whereas the Adjacent Reciprocal Effects Model produced two such correlation residuals), the correlation residual between anger at time 3 and worry at time 3 were close to the .15 cut-off for both models. The remaining correlation residuals were either across-time relationships between worry and anger (both negative positive emotions) or across-time relationships between enjoyment, pride, and boredom (all activity emotions). Thus, in the future, the shared variance among these emotions may need to be accounted by estimating relationships between these emotions.

In sum, both the Adjacent Reciprocal Effects Model and the Complete Reciprocal Effects Model fit the data approximately the same. Furthermore, both of these models fit the data well in an absolute sense. Thus, I concluded that both models were plausible explanations for the variable relations.

**Interpretation of Direct and Indirect Effects**

Given both, the Adjacent Reciprocal Effects Model and the Complete Reciprocal Effects Model fit the data reasonably well, the parameter estimates from each model were interpreted. The direct and indirect effects from the Adjacent Reciprocal Effects Model are summarized in Table 13-15. The direct and indirect effects from the Complete Reciprocal Effects model are summarized in Table 16-18.

Effects that were statistically significant in a simpler model can become non-significant in a more complex model because the effects are tested in the presence of
greater number of other effects. For example, the effect of boredom at time 1 on perceived value at time 2 was statistically significant in the Adjacent Reciprocal Effects Model but was not statistically significant in Complete Reciprocal Effects Model. The opposite can also happen: an effect may not be significant in a simpler model but can become significant in a more complex model. For example, the effect of perceived value at time 1 on pride at time 2 was not statistically significant in the Adjacent Reciprocal Effects Model but was statistically significant in Complete Reciprocal Effects Model. Table 19 displays the 14 direct effects (out of 28 estimated direct effects) that changed in terms of statistical significance across the two models.

The effects from the Complete Reciprocal Effects Model were interpreted for two reasons. First, in this more complex model, the effect of one construct on another construct was evaluated when controlling for a greater number of other effects. Second, one of the goals of the current study was to examine effects specified by EV theory, in the presence of the effects specified by the CV theory and vice versa. Since the reciprocal effects models were built on CV theory, only the Complete Reciprocal Effects Model estimated the effects of EV theory. Thus, to examine EV theory effects in the presence of CV theory effects, the direct and indirect effects from the Complete Reciprocal Effects Model were interpreted below.

Following the interpretation of statistically significant direct effects, stationarity for each of the statistically significant direct effects was examined. Stationarity refers to the unchanging causal structure of constructs over time (Cole & Maxwell, 2003, Preacher, 2015). To test stationarity, a model constraining the direct effect of one construct on subsequent construct was constrained to be equal at both times lags. For
example, to test whether the autoregressive effect of perceived value on subsequent perceived value was equal over time, the effect of perceived value at time 1 on perceived value at time 2 was constrained to equal the effect of perceived value at time 2 on perceived value at time 3. Thus, a one degree of freedom test can be performed to evaluate whether the statistically significant direct effects are equal over time.

**Direct effects.** Perceived value had strong positive autoregressive effects on subsequent perceived value from time 1 to time 2 (.831) and from time 2 to time 3 (.813). This effect was stationary over time. Perceived value had negative direct effects on subsequent anger from time 1 to time 2 (-.052) and from time 2 to time 3 (-.053). This effect was stationary over time. At time 1, perceived value had a positive direct effect on enjoyment (.040) at time 2 and a positive direct effect on pride (.029) at time 2, but neither of these two effects were statistically significant at the later time lag (i.e., time 2 to time 3). However, the model constraining these effects to be equal over time was not statistically different from the baseline model. Thus, both effects were stationary over time. Perceived value at time 1 did not have a statistically significant effect on boredom at time 2, but perceived value at time 2 did have a significant effect on boredom at time 3 (-.044). This effect was stationary over time. Perceived value did not have a statistically significant effect on subsequent worry, examinee effort, or test performance at either time lags. In sum, stationarity was observed for all the statistically significant direct effects of perceived value on subsequent constructs.

Anger had strong positive autoregressive effects on subsequent anger from time 1 to time 2 (.808) and from time 2 to time 3 (.815). The autoregressive effect was stationary over time. At time 1, anger had negative effects on perceived value at time 2 (-.092),
examinee effort at time 2 (-.067), and test performance at time 2 (-.121). However, anger at time 2 did not have statistically significant effects on perceived value, examinee effort or test performance at time 3. Only the effect of anger on subsequent perceived value was not equal over time: $\chi^2 (1) = 5.87, p = .0154$. Stationarity was established for the direct effects of anger on subsequent anger, effort, and test performance.

Boredom had strong positive autoregressive effects on subsequent boredom from time 1 to time 2 (.800) and from time 2 to time 3 (.842). Boredom had a negative direct effect on subsequent examinee effort from time 1 to time 2 (-.154) and from time 2 to time 3 (-.101). Boredom at time 1 did not have a statistically significant effect on test performance at time 2; however, boredom at time 2 had a positive direct effect on test performance at time 3 (.103). Stationarity for all these direct effects was supported. Lastly, boredom did not have a statistically significant effect on subsequent perceived value at either time lags.

Enjoyment had strong positive autoregressive effects on subsequent enjoyment from time 1 to time 2 (.756) and from time 2 to time 3 (.844). Stationarity for this effect was not established: $\chi^2 (1) = 10.09, p = .001$. Enjoyment did not have statistically significant effects on subsequent perceived value, examinee effort, or test performance at either time lags.

Pride had strong positive autoregressive effects on subsequent pride from time 1 to time 2 (.829) and from time 2 to time 3 (.860). Pride had a positive direct effect on subsequent examinee effort from time 1 to time 2 (.159) and from time 2 to time 3 (.119). Pride at time 1 had a direct positive effect on perceived value at time 2 (.075), but pride at time 2 did not have a direct effect on perceived value at time 3. Stationarity for all direct
effects was supported. Pride did not have a statistically significant effect on subsequent test performance at either time lags.

Worry had strong positive autoregressive effects on subsequent worry from time 1 to time 2 (.799) and from time 2 to time 3 (.849). However, stationarity was not established the autoregressive effect: $\chi^2 (1) = 4.26, p = .039$. Worry had a positive effect on subsequent perceived value from time 1 to time 2 (.100) and from time 2 to time 3 (.069) and on subsequent examinee effort from time 1 to time 2 (.073) and from time 2 to time 3 (.053). Worry at time 1 did not have an effect on test performance at time 2 but worry at time 2 had a negative effect on test performance time 3 (-.123). Stationarity was established for the direct effects of worry on subsequent perceived value, effort, and test performance, but not for the effect of worry on subsequent worry.

Examinee effort had strong positive autoregressive effects on subsequent examinee effort from time 1 to time 2 (.727) and from time 2 to time 3 (.813). The autoregressive effect was not stationary over time: $\chi^2 (1) = 6.55, p = .010$. Examinee effort had a positive effect on subsequent pride from time 1 to time 2 (.065) and from time 2 to time 3 (.059). Examinee effort had a positive effect on subsequent test performance from time 1 to time 2 (.194) and from time 2 to time 3 (.223). Examinee effort had a negative effect on subsequent anger from time 1 to time 2 (-.075) and from time 2 to time 3 (-.039). Examinee effort at time 1 did not have an effect on perceived value at time 2, but examinee effort at time 2 had a positive direct effect on perceived value at time 3 (.089). Examinee effort at time 1 had a negative effect on boredom at time 2 (-.066) and worry at time 2 (-.053), but neither of these effects were statistically significant at the later time lag. Thus, of the eight tests for stationarity, only the effect of
effort on subsequent effort differed over time. Thus, stationarity was established for the effects of effort on subsequent value, anger, boredom, pride, worry, and test performance, but not for the autoregressive effect of effort on subsequent effort. Lastly, examinee effort did not have an effect on subsequent enjoyment at either of the time lags.

Test performance had strong positive autoregressive effects on subsequent test performance from time 1 to time 2 (.584) and from time 2 to time 3 (.555). The autoregressive effect was stationary over time. Test performance had a positive effect on subsequent examinee effort from time 1 to time 2 (.048) and from time 2 to time 3 (.036). The effect of test performance on subsequent examinee effort was not stationary over time: $\chi^2(1) = 6.64, p = .010$. Test performance at time 1 had a negative effect on worry at time 2 (-.042) but test performance at time 2 did not have an effect on worry at time 3. This effect of test performance on subsequent worry was not stationary over time: $\chi^2(1) = 5.68, p = .017$. Test performance did not have a direct effect on perceived value, anger, boredom, enjoyment, and pride at either of the time lags.

**Indirect Effects.**

*Indirect effects of perceived value.* Perceived value at time 1 had a strong positive indirect effect on perceived value at time 3 (.683). Perceived value at time 1 also had positive indirect effects on enjoyment at time 3 (.042), pride at time 3 (.036), and effort at time 3 (.061) and a negative indirect effect on anger at time 3 (-.088). Lastly, perceived value at time 1 did not have statistically significant indirect effects on boredom, worry, or test performance at time 3.

The total indirect effects reported above can be broken down into indirect effects via multiple mediators. That is, perceived value at time 1 exerted its effect on perceived
value at time 3 via all eight constructs at time 2. Thus, eight indirect effects of perceived value at time 1 on perceived value at time 3 were calculated and tested for statistical significance using the distribution of product method (Tofighi & McKinnon, 2011). Of the eight indirect effects, only the indirect effect of perceived value at time 1 on perceived value at time 3 via perceived value at time 2 was statistically significant (estimate = .676, standard error = .024, Z = 28.350).

In fact, in many cases when examining the total indirect effects, the only statistically significant individual effect involves the same construct measured at two different times. For example, the effect of perceived value at time 1 on anger at time 3 was mediated through four constructs at time 2: perceived value, anger, examinee effort, and test performance. Two of these indirect effects were statistically significant: the indirect effect via perceived value at time 2 (estimate = -.044, s.e. = 0.013, Z = -3.300) and the indirect effect via anger at time 2 (estimate = -.042, s.e. = 0.016, Z = -2.731). Thus, in this situation there are two individual indirect effects, however both of these effects involve the same two constructs (the effect of perceived value at time 1 on anger at time 3 via perceived value at time 2 or the effect of perceived value at time 1 on anger at time 3 via anger at time 2).

The effect of perceived value at time 1 on enjoyment at time 3 was mediated through four constructs at time 2: perceived value, enjoyment, examinee effort, and test performance. Of the four indirect effects, only the indirect effect through enjoyment at time 2 was statistically significant (estimate = .034, s.e. = 0.014, Z = 2.349).

The effects of perceived value at time 1 on pride and effort at time 3 were mediated through four and eight constructs at time 2 respectively. In both cases, the total
Indirect effects were statistically significant (as noted above). However, when examined individually, none of the indirect effects reached statistical significance.

**Indirect effects of anger.** Anger at time 1 had a strong positive indirect effect on anger at time 3 (.666). Anger at time 1 also had negative indirect effects on perceived value at time 3 (-.081) and effort at time (-.063). Anger at time 1 did not have statistically significant indirect effects on boredom, enjoyment, pride, worry, or test performance at time 3.

The effect of anger at time 1 on anger at time 3 was mediated via four constructs at time 2: perceived value, anger, examinee effort, and test performance. Two of these indirect effects were statistically significant: the effect via perceived value at time 2 (estimate = .005, s.e. = .002, Z = 2.28) and the effect via anger at time 2 (estimate = .0659, s.e. = .024, Z = 27.894).

The effect of anger at time 1 on perceived value at time 3 was mediated via four constructs at time 2: perceived value, anger, examinee effort, and test performance. Only the indirect effect via perceived value at time 2 was statistically significant (estimate = -.075, s.e. = .023, -3.274).

The effect of anger at time 1 on examinee effort at time 3 was mediated via four constructs at time 2: perceived value, anger, examinee effort, and test performance. Only the indirect effect via examinee effort at time 2 was statistically significant (estimate = -.054, s.e. = .024, Z = -2.229).

**Indirect effects of boredom.** Boredom at time 1 had a strong positive indirect effect on boredom at time 3 (.679) and positive indirect effect on anger at time 3 (.009). Boredom at time 1 also had negative indirect effects on perceived value at time 3 (-.077),
enjoyment at time 3 (-.005), pride at time 3 (-.010), and effort (-.206). Boredom at time 1 did not have statistically significant effects on worry or test performance at time 3.

The effect of boredom at time 1 on perceived value at time 3 was mediated via four constructs at time 2: perceived value, anger, examinee effort, and test performance. Only the indirect effect via examinee effort at time 2 was statistically significant (estimate = -.014, s.e. = .004, Z = -3.141).

The effect of boredom at time 1 on anger at time 3 was mediated via three constructs at time 2: perceived value, examinee effort, and test performance. Only the indirect effect via examinee effort at time 2 was statistically significant (estimate = .006, s.e. = .003, Z = 2.169).

The effect of boredom at time 1 on boredom at time 3 was mediated via four constructs at time 2: perceived value, boredom, examinee effort, and test performance. Only the indirect effect via boredom at time 2 statistically significant (estimate = -.009, s.e. = .003, Z = -3.221).

The effect of boredom at time 1 on pride at time 3 was mediated via three constructs at time 2: perceived value, examinee effort, and test performance. Only the indirect effect via examinee effort at time 2 was statistically significant (estimate = -.009, s.e. = .003, Z = -3.221).

The effect of boredom at time 1 on examinee effort at time 3 was mediated via four constructs at time 2: perceived value, boredom, examinee effort, and test performance. Two of these indirect effects statistically significant: the effect via boredom at time 2 (estimate = -.081, s.e. = .019, Z = -4.181) and the effect via examinee effort at time 2 (estimate = -.125, s.e. = .025, Z = -5.087).
The effect of boredom at time 1 on enjoyment at time 3 was mediated via three constructs at time 2: perceived value, examinee effort, and test performance. However, none of these indirect effects were statistically significant individually.

**Indirect effects of enjoyment.** Enjoyment at time 1 had a strong positive indirect effect on enjoyment at time 3 (.638). Enjoyment at time also had a negative indirect effect on effort at time 3 (-.075). Enjoyment at time 1 did not have statistically significant indirect effect on perceived value, anger, boredom, pride, worry, or test performance at time 3.

The effect of enjoyment at time 1 on enjoyment at time 3 was mediated via four constructs at time 2: perceived value, enjoyment, examinee effort, and test performance. Only the indirect effect via enjoyment at time 2 was statistically significant (estimate = .637, s.e. = .023, Z = 28.122).

The effect of enjoyment at time 1 on examinee effort at time 3 was mediated via four constructs at time 2: perceived value, enjoyment, examinee effort, and test performance. However, none of these indirect effects were statistically significant individually.

**Indirect effects of pride.** Pride at time 1 had a strong positive indirect effect on pride at time 3 (.724), and positive indirect effects on perceived value at time 3 (.112) and effort at time 3 (.231). Pride at time 1 also had a negative indirect effect on anger at time 3 (-.01). Pride at time 1 did not have statistically significant indirect effects on boredom, enjoyment, worry, or test performance at time 3.

The effect of pride at time 1 on pride at time 3 was mediated via four constructs at time 2: perceived value, pride, examinee effort, and test performance. Two of these
indirect effects were statistically significant: the indirect effect via pride at time 2
(estimate = .713, s.e. = .025, Z = 28.336) and the indirect effect via worry at time 2
(estimate .009, s.e. = .003, Z = 2.888).

The effect of pride at time 1 on perceived value at time 3 was mediated via four
constructs at time 2: perceived value, pride, examinee effort, and test performance. Two
of these indirect effects were statistically significant: the indirect effect via perceived
value at time 2 (estimate = .061, s.e. = .029, Z = 2.139) and the indirect effect via
examinee effort at time 2 (estimate .014, s.e. = .005, Z = 2.829).

The effect of pride at time 1 on examinee effort at time 3 was mediated via four
constructs at time 2: perceived value, pride, examinee effort, and test performance. Two
of these indirect effects were statistically significant: the indirect effect via pride at time 2
(estimate = .099, s.e. = .025, Z = 3.943) and the indirect effect via worry at time 2
(estimate .129 s.e. = .032, Z = 4.053).

The effect of pride at time 1 on anger at time 3 was mediated via three constructs
at time 2: perceived value, examinee effort, and test performance. Only the indirect effect
via examinee effort at time 2 was statistically significant (estimate = -.006, s.e. = .003 , Z
= -2.047).

**Indirect effects of worry.** Worry at time 1 had a strong indirect effect on worry at
time 3 (.676), and positive indirect effects on perceived value at time 3 (.143), pride at
time 3 (.006), and effort at time 3 (.106). Worry at time 1 had negative indirect effects on
anger at time 3 (-.008), boredom at time (-.005), and test performance at time 3 (-.083).
Worry at time 1 did not have a statistically significant indirect effect on enjoyment at
time 3.
The effect of worry at time 1 on perceived value at time 3 was mediated via four constructs at time 2: perceived value, examinee effort, and test performance. Three of these indirect effects were statistically significant: the indirect effect via perceived value at time 2 (estimate = .081, s.e. = .019, Z = 4.322), the indirect effect via worry at time 2 (estimate = .055, s.e. = .018, Z = 3.127), and the indirect effect via examinee effort (estimate = .006, s.e. = .003, Z = 2.260).

The effect of worry at time 1 on anger at time 3 was mediated via three constructs at time 2: perceived value, examinee effort, and test performance. Only the indirect effect via perceived value at time 2 was statistically significant (estimate = -.005, s.e. = .002, Z = -2.592).

The effect of worry at time 1 on boredom at time 3 was mediated via three constructs at time 2: perceived value, examinee effort, and test performance. Only the indirect effect via perceived value at time 2 was statistically significant (estimate = -.004, s.e. = .002, Z = -2.182).

The effect of worry at time 1 on pride at time 3 was mediated via three constructs at time 2: perceived value, examinee effort, and test performance. Only the indirect effect via examinee effort at time 2 was statistically significant (estimate = .004, s.e. = .002, Z = 2.292).

The effect of worry at time 1 on worry at time 3 was mediated via four constructs at time 2: perceived value, examinee effort, and test performance. Only the indirect effect via worry at time 2 was statistically significant at time 2 (estimate = .678, s.e. = .020, Z = 33.752).
The effects of worry at time 1 on examinee effort at time 3 was mediated via four constructs at time 2: perceived value, examinee effort, and test performance. Two of these indirect effects were statistically significant: the indirect effect via worry at time 2 (estimate = .042, s.e. = .017, Z = 2.519) and the indirect effect via examinee effort at time 2 (estimate = .059, s.e. = .021, Z = 2.799).

The effects of worry at time 1 on test performance at time 3 was mediated via four constructs at time 2: perceived value, examinee effort, and test performance. Two of these indirect effects were statistically significant: the indirect effect via worry at time 2 (estimate = -.098, s.e. = .029, Z = -3.404) and the indirect effect via examinee effort at time 2 (estimate = .016, s.e. = .006, Z = 2.530).

**Indirect effects of examinee effort.** Effort at time 1 had a strong positive indirect effect on effort at time 3 (.611) and positive indirect effects on perceived value at time 3 (.093), pride at time 3 (.098), and test performance at time 3 (.263). Effort at time 1 also had negative indirect effects on anger at time (-.090), boredom at time 3 (-.067), and worry at time 3 (-.062). Effort at time 1 did not have statistically significant indirect effect on enjoyment at time 3.

The effect of examinee effort at time 1 on perceived value at time 3 was mediated via eight constructs at time 2: perceived value, anger, boredom, enjoyment, pride, worry, examinee effort, and test performance. Two of these indirect effects were statistically significant: the indirect effect via worry at time 2 (estimate = -.004, s.e. = .002, Z = -2.091) and the indirect effect via examinee effort at time 2 (estimate = .065, s.e. = .016, Z = 4.005).
The effect of examinee effort at time 1 on anger at time 3 was mediated via four constructs at time 2: perceived value, anger, examinee effort, and test performance. Two of these indirect effects were statistically significant: the indirect effect via anger at time 2 (estimate = -0.061, s.e. = 0.016, Z = -3.930) and the indirect effect via examinee effort at time 2 (estimate = -0.028, s.e. = 0.012, Z = 2.428).

The effect of examinee effort at time 1 on boredom at time 3 was mediated via four constructs at time 2: perceived value, anger, examinee effort, and test performance. Only the indirect effect via boredom at time 2 was statistically significant (estimate = -0.056, s.e. = 0.018, Z = -3.133).

The effect of examinee effort at time 1 on pride at time 3 was mediated via four constructs at time 2: perceived value, anger, examinee effort, and test performance. Two of these indirect effects were statistically significant: the indirect effect via pride at time 2 (estimate = 0.056, s.e. = 0.013, Z = 4.310) and the indirect effect via examinee effort at time 2 (estimate = 0.043, s.e. = 0.010, Z = 4.168).

The effect of examinee effort at time 1 on worry at time 3 was mediated via four constructs at time 2: perceived value, anger, examinee effort, and test performance. Only the indirect effect via worry at time 2 was statistically significant (estimate = -0.045, s.e. = 0.015, Z = -2.940).

The effect of examinee effort at time 1 on examinee effort at time 3 was mediated via eight constructs at time 2: perceived value, anger, boredom, enjoyment, pride, worry, examinee effort, and test performance. Four of these indirect effects were statistically...
significant: the indirect effect via boredom at time 2 (estimate = .007, s.e. = .003, Z = 2.474), the indirect effect via pride at time 2 (estimate = .008, s.e. = .003, Z = 2.884), the indirect via examinee effort at time 2 (estimate = .591, s.e. = .025, Z = 23.246), and the indirect effect via test performance at time 2 (estimate = .007, s.e. = .003, Z = 2.115).

The effect of examinee effort at time 1 on test performance at time 3 was mediated via eight constructs at time 2: perceived value, anger, boredom, enjoyment, pride, worry, examinee effort, and test performance. Three of these indirect effects were statistically significant: the indirect effect via worry at time 2 (estimate = .007, s.e. = .003, Z = 2.178), the indirect effect via examinee effort at time 2 (estimate = .172, s.e. = .027, Z = 6.055), and the indirect via examinee test performance at time 2 (estimate = .108, s.e. = .023, Z = 4.734).

**Indirect effects of test performance.** Test performance at time 1 had positive indirect effects on test performance at time 3 (.342) and effort at time 3 (.055). Test performance at time 1 also had a negative indirect effect on worry at time 3 (-.034). Test performance at time 1 did not have statistically significant indirect effects on perceived value, anger, boredom, enjoyment, or pride at time 3.

The effect of test performance at time 1 on worry at time 3 was mediated via four constructs at time 2: perceived value, worry, examinee effort, and test performance. Only the indirect effect via worry was statistically significant (estimate = -.036, s.e. = .014, Z = -2.622).

The effect of test performance at time 1 on examinee effort at time 3 was mediated via eight constructs at time 2: perceived value, anger, boredom, enjoyment, pride, worry, examinee effort, and test performance. Two of these indirect effects were
statistically significant: the indirect effect via examinee effort at time 2 (estimate = .039, s.e. = .017, Z = 2.281) and the indirect effect via test performance at time 2 (estimate = .021, s.e. = .009, Z = 2.375).

The effect of test performance at time 1 on test performance at time 3 was mediated via eight constructs at time 2: perceived value, anger, boredom, enjoyment, pride, worry, examinee effort, and test performance. Three of these indirect effects were statistically significant: the indirect effect via worry at time 2 (estimate = .005 , s.e. = .003, Z = 2.028), the indirect effect via examinee effort at time 2 (estimate = .011, s.e. = .005, Z = 2.120), and the indirect effect via test performance (estimate = .324, s.e. = .023, Z = 13.830).
Chapter 5: Discussion

The purpose of the current study was two-fold. First, the EV theory hypothesis, which states that perceived value of a test influences test performance via examinee effort, was evaluated using longitudinal data. Second, using CV theory as a framework, the effects of five test emotions on examinee effort and test performance were examined. The results from the current study are concisely summarized and their implications are presented below. Lastly, limitations of the current study and recommendations for future studies are discussed.

Emotions not perceived test importance predicted effort and performance

When evaluated using longitudinal data, the EV theory hypothesis that perceived value influences effort which subsequently influences test performance was not supported. In the current study, examinee effort had a direct effect on subsequent test performance at both time lags and was the strongest predictor of subsequent test performance at both time lags. The observed magnitude of this unstandardized direct effect (.194 & .233, at the first and second time lag, respectively) was lower than the observed effect of effort on math test in a previous study (Cole et al., 2008). However, in the study by Cole and colleagues (2008), researchers did not control for previous levels of test performance or examinee effort; thus, the effects in their study may be inflated. In addition to examinee effort, three test emotions had direct effects on subsequent test performance. At time 1, anger had a negative effect on test performance at time 2, whereas at time 2 boredom and worry had positive and negative effects on test performance at time 3, respectively. The magnitude of the effects of test emotions on subsequent test performance was lower than the magnitude of the effects of examinee
effort and lower than observed effects of test emotions on test performance in a previous cross-sectional study (Pekrun et al., 2002).

The non-significant indirect effect of perceived value on test performance via examinee effort, as hypothesized in EV theory, can be explained by the lack of direct effect of perceived value on subsequent effort. That is, perceived value did not have a direct effect on subsequent examinee effort at either time lag, when controlling for other constructs in the study. These findings align with a longitudinal study that modeled perceived value, anger, pride, and examinee effort during a low-stakes test at three time points (Finney et al., 2018). It could be argued that inclusion of strong autoregressive effects when predicting subsequent examinee effort does not leave any variance for perceived value to explain. However, other constructs (reviewed below) in the current study had effects on subsequent examinee effort, in the presence of perceived value and the autoregressive effects. Thus, the current study suggests that perceived value of a test is not a direct antecedent of examinee effort, as hypothesized in EV theory, after controlling for other variables. One implication of this finding for practitioners is that trying to manipulate perceived test value may be in vain, as perceived value may not have an effect on subsequent examinee effort in low-stakes testing contexts. Instead, researchers should examine the role of other constructs on examinee effort, such as emotions.

According to CV theory, the direct antecedents of students’ motivation are emotions. The current study provides evidence that test-taking emotions have direct and indirect effects on examinee effort. As expected, anger and boredom had negative direct effects on subsequent examinee effort. These findings align with results from previous
studies situated in CV theory, in which anger and boredom had negative relationships with students’ effort (Pekrun et al., 2002; Pekrun et al., 2004). Pride and worry had positive direct effects on subsequent examinee effort. The positive effects of pride are expected given that in CV theory, positive activating emotions are hypothesized to affect effort positively (Pekrun, 2017). However, the positive effects of worry on subsequent examinee effort were not expected, and in fact, are in the opposite direction of what the previous studies have found (Pekrun et al., 2002; Pekrun et al., 2004). The magnitude of the effects of test emotions on subsequent effort was small (they ranged from -.067 to .159) and these effects were lower than observed in previous cross-sectional studies.

Lastly, in the current study, test-taking enjoyment did not have an effect on subsequent examinee effort. Thus, the unexpected results of the effects (or lack of effects) of worry and enjoyment on subsequent examinee effort only partially support CV theory in low-stakes testing environment.

Anger, boredom, pride, and worry after the first third of the test had indirect effects on students’ effort at the end of the test. For anger and pride, these indirect effects were mediated via students’ effort after the second third of the test. Boredom and worry after the first third of the test had two indirect effects on students’ effort at the end of the test. For boredom, the effect was mediated via boredom and students’ effort after the second third of the test. For worry, the effect was mediated via worry and students’ effort after the second third of the test. Thus, three distinct variable mediation was not observed in the indirect effects summarized above. All the indirect effects of test emotions on subsequent examinee effort were mediated either via examinee effort or the same emotion that started the effect. In other words, there was always one autoregressive effect
involved. The implication from these indirect effects is that anger, boredom, pride, and worry have long lasting effects on students’ effort. Importantly, the magnitude of the indirect effects ranged from -.081 to .129 on a 1 to 5 scale; thus, the indirect effects were rather small.

Two total indirect effects on test performance are worth noting. First, worry had indirect effects on test performance via two mediators: worry and examinee effort. The indirect effect via worry involved an autoregressive effect, and thus the implication of this effect is no different than the implication of indirect effects summarized in the previous paragraph. In contrast, the indirect effect of worry on test performance via examinee effort was a three distinct variable mediated effect. This effect aligns with CV theory, in which examinee effort is hypothesized to mediate the effects of test emotions on test performance. Of all the emotions, only worry had an indirect effect on test performance. Thus, the current study provides only minimal support for the indirect effects of test emotions on performance via motivation, as hypothesized in CV theory.

Second, students’ effort had a total indirect effect on test performance via three mediators. Two of the statistically significant indirect effects involved either the autoregressive effect of effort at the earlier time lag (i.e., time 1 to time 2), or the autoregressive effect of test performance at the later time lag (i.e., time 2 to time 3). The third indirect effect was mediated via worry at time 2.

To sum up, worry had an indirect effect on test performance via effort, whereas effort had an indirect effect on test performance via worry. Both of these effects were small (.016 & .007, respectively). Recall, worry and examinee effort also had direct effects on subsequent test performance. The direct effect of examinee effort on
subsequent test performance was positive and has been empirically supported in previous studies in low-stakes tests (Cole et al., 2008). The direct effect of worry on subsequent test performance was negative and has been empirically supported in previous studies (Pekrun et al., 2004), but the indirect effect was positive. It appears that worry and examinee effort are both important constructs that affect test performance in low-stakes testing environments.

Contrary to the CV theory predictions, perceived value did not have indirect effects on students’ effort via test emotions. The total indirect effect of perceived value measured early in the test on effort later on the test was statistically significant (.061). However, in the current study, the total indirect effect of perceived value on effort was composed of indirect effects via eight different constructs at time 2. When evaluated individually, none of the eight direct effects had a statistically significant effect on subsequent examinee effort.

Statistical power of the indirect effect could have been an issue for some effects noted above. For example, perceived value at time 1 had an effect on pride at time 2 (.029), and pride at time had an effect on examinee effort at time 3 (.119). Thus, both effects that were involved in the indirect effect of perceived value on examinee effort via pride were statistically significant, but the indirect effect was not. The effects of perceived value at time 1 on pride and other emotions at time 2 were small in magnitude; thus, the statistical tests for these small indirect effects may have been underpowered.

The lack of effect of perceived value on emotions, effort or performance could be due to the type of value modeled. Putwain and colleagues (2018) conducted a study in which the effects of three types of perceived value on enjoyment and boredom were
estimated. The authors found that intrinsic, achievement, and utility value had an effect on students’ subsequent enjoyment, whereas only intrinsic value related to subsequent boredom. The effects of intrinsic value on subsequent enjoyment (.31) and boredom (-.51) were greatest in magnitude. The effect of achievement value (which closely align with perceived value measure used in the current study) on subsequent enjoyment was greater in magnitude in the Putwain and colleagues study (.11) than in the current study (.04), which could be due to their lack of autoregressive effects. Moreover, in the Putwain et al. (2018) study, the multiplicative effect of achievement value and perceived control was statistically significant; thus, the direct effect of achievement value on subsequent enjoyment should be interpreted with the interaction term in mind. Nonetheless, the study by Putwain and colleagues (2018) illustrates that the effects of perceived value on subsequent test emotions may depend on the type of perceived value measured. Related, Cole and colleagues (2008) found that different values (interest, usefulness, importance) predicted effort differentially on four different subject tests (English, Math, Science, and Social Studies). Usefulness and importance had positive effects on effort on all of the four tests, whereas interest had a negative effect on effort only on English test. The magnitude of effects of importance (which closely align with perceived value measure used in the current study) on effort ranged from .32 to .43. However, the magnitude of these effects may be due to lack of autoregressive effects. Future studies should include other types of perceived value measures and examine whether effects of perceived value on test emotions or the indirect effects of perceived value on examinee effort via test emotions would emerge.
Reciprocal Relations

Previous studies have demonstrated the reciprocal relations between students’ emotions and achievement (Gibbons et al., 2018; Pekrun et al., 2014; Putwain et al., 2018). Most of these studies measured class-related or learning-related emotions and found positive reciprocal relations between positive emotions (e.g., pride, enjoyment) and achievement and negative reciprocal relations between negative emotions (e.g., boredom, anxiety) and achievement. For example, boredom had an effect on subsequent test performance, which then had an effect on subsequent boredom (Pekrun et al., 2014). Importantly, the study by Reeve and colleagues (2014) replicated the reciprocal relations with test-related emotions and achievement. They found that positive emotions had positive reciprocal relations with achievement (e.g., positive emotions had an effect on subsequent test performance, which then had an effect on subsequent positive emotions). However, in their study, pride and relief were combined to represent positive emotions composite, whereas anger and shame were combined to represent negative emotions composite. In the current study, when examined individually, only worry had a negative reciprocal relation with test performance (e.g., performance at time 1 had an effect on worry at time 2 (-.042), which then had an effect on performance at time 3 (-.123)).

Thus, in the current study, the reciprocal relation between test emotions and test performance were supported only for worry and test performance. In the basic CV theory model (Figure 2), the effects of test emotions on achievement are hypothesized to be mediated via examinee effort. Thus, it is possible that the reciprocal relations may also be mediated via examinee effort. However, as noted in chapter 2, the research on reciprocal
effects in CV theory has only recently began. More studies examining reciprocal relations are needed.

Test performance and examinee effort also had a positive reciprocal relation over time. It is worth noting the difference between reciprocal relation and reciprocal effect. If the construct hypothesized to be later in the CV theory causal chain (e.g., test performance) has an effect on a construct hypothesized to be earlier in the causal chain (e.g., examinee effort), which then has an effect on the construct hypothesized to be later in the causal chain (e.g., test performance), then this is termed a reciprocal relation. In other words, the reciprocal relation occurs only if the following two effects are observed. First, the reciprocal effect of a construct hypothesized to be later in the causal chain (e.g., test performance) on the construct hypothesized to be earlier (e.g., examinee effort) in the causal chain. Second, the direct effect of the construct hypothesized to be earlier in the causal chain (e.g., examinee effort) on the construct hypothesized to be later in the causal chain (e.g., test performance). For example, test performance at time 1 had a reciprocal effect on examinee effort at time 2 (.048), which then had a direct effect on test performance at time 3 (.223). Thus, overall, these two effects form a reciprocal relation between test performance and examinee effort. Test performance at time 2 also had a reciprocal effect on examinee effort at time 3 (.036). However, since the constructs were measured only at three times, it is impossible to determine if examinee effort at time 3 would have a direct effect on subsequent test performance.

Currently, no studies framed in CV theory have examined these effects, thus the current study provides evidence for the theoretical reciprocal relation between test performance and examinee effort (Pekrun, 2017). No studies have examined reciprocal
relations between test performance and perceived value. In the current study, these relations were not observed.

Next, the reciprocal relations between examinee effort and test emotions were evaluated. According to CV theory, positive emotions should have positive reciprocal relations with examinee effort, whereas negative emotions should have negative reciprocal relations with examinee effort (Pekrun, 2017). In the current study, examinee effort had reciprocal relations with boredom, worry, and pride. That is, examinee effort at time 1 had a negative effect on boredom at time 2 (-.066), which then had a negative effect on examinee effort at time 3 (-.101). Examinee effort at time 1 had a negative effect on worry at time 2 (-.053), which then had a positive effect on examinee effort at time 3 (.053). Lastly, examinee effort at time 1 had a positive effect on pride at time 2 (.065), which then had a positive effect on examinee effort at time 3 (.119). Two other reciprocal effects were observed: examinee effort at time 2 had a positive reciprocal effect on pride at time 3 (.059) and examinee effort at time 2 had a negative reciprocal effect on anger at time 3 (-.039). The current study serves as the first study to test the reciprocal relations between examinee effort and test emotions, as specified in CV theory.

According to CV theory, examinee effort and perceived value should have positive reciprocal relations over time. However, in the current study, evidence for the reciprocal relations between examinee effort and perceived value were not observed but the reciprocal effect of examinee effort at time 2 on perceived value at time 3 (.089) was found.

Lastly, the reciprocal relations between perceived value and test emotions were evaluated. According to CV theory, positive emotions such as pride or enjoyment should
have positive reciprocal relations with perceived value (e.g., pride should have a positive effect on subsequent perceived value, which then should have a positive effect on subsequent pride). Negative emotions such as anger or boredom should have negative reciprocal relations with perceived value (e.g., anger should have a negative effect on subsequent perceived value, which then should have a negative effect on subsequent anger) (Pekrun, 2017). Only one study empirically examined these relations over time. Putwain and colleagues (2018) found that learning-related enjoyment had a positive effect on subsequent intrinsic and utility value and then intrinsic and utility value had positive effects on subsequent enjoyment. They also found positive reciprocal relation for achievement value and enjoyment; however, the effect of an interaction between perceived control and achievement value on subsequent enjoyment was observed. Moreover, boredom had a negative reciprocal relation with intrinsic value and reciprocal effects of boredom on subsequent achievement and utility value.

In the current study, reciprocal relation between enjoyment and perceived value or reciprocal relation between boredom and perceived value were not observed. Only anger and perceived value had negative reciprocal relation: anger at time 1 had a negative reciprocal effect on perceived value at time 2 (-.092), which then had a negative effect on perceived value (-.053). Pride only at time 1 had a positive reciprocal effect on perceived value at time 2 (.075). Worry had positive reciprocal effects on subsequent perceived value at both time lags. That is, worry at time 1 had a positive reciprocal effect on perceived value at time 2 (.100), whereas worry at time 2 had a positive reciprocal effect on perceived value at time 3 (.069). Thus, the current study provides partial support for the reciprocal relations between test emotions and perceived value over time.
**Limitations and Future Studies**

The current study has several limitations. First, I was not able to test any of the proposed models that involved self-efficacy or value at the beginning of the test due to poor measurement. The results from the current study should be interpreted cautiously, as the estimated models were misspecified. Moreover, researchers have argued (see Flake et al., 2015; Nagengast et al., 2011) for modeling the interaction between self-efficacy and perceived value. Due to poor measurement in the current study, I was not able to examine the interaction effects. Future studies should address this limitation by administering a higher quality instruments of self-efficacy and perceived value and modeling the effects of the interaction.

Second, the study uncovered many effects among perceived value, test emotions, examinee effort, and test performance. However, it could be argued that the practical significance of these effects is minimal. The highest unstandardized effects observed in the study were the autoregressive effects of constructs close in time. For example, the autoregressive effects of perceived value at time 1 on perceived value at time 2 was .83. On 1 to 5 point Likert scale, such effects are undoubtedly important to consider. However, many other direct and indirect effects were not high. In fact, the majority of them were below |.10|. One of the reasons why many observed effects were small in magnitude is due to high autoregressive effects. However, if the goal is to infer change of constructs over time, autoregressive effects are important to include in the model. The best predictor of any construct is often the earlier measure of the same construct. If the autoregressive effects are not included in the model, then other effects will appear larger.
that they should and may seem practically important; however, such findings should not be trusted.

Third, in the current study, only five test emotions and their effects on perceived value, examinee effort, and test performance were evaluated. According to CV theory, students experience other emotions in achievement situations (e.g., gratitude, relief, shame, and frustration). Data in the present study were collected in an operational testing setting, thus I was limited with the number of emotions I could measure. Anger, boredom, and worry have received the most attention in the CV theory literature (Pekrun, 2017). Future studies should examine other emotions. Moreover, CV theory provides a comprehensive set of predictions that involve many other constructs I did not address in the current study. Constructs that fall into the “environment” category of CV theory (Figure 2), were not addressed in the current study at all.

Lastly, the method in which longitudinal data were collected may have altered students’ experience of taking the test. That is, students were asked to complete the same measures of perceived value, effort, and test emotions three times during a relatively short amount of time. I was one of the proctors during the Assessment Day, and I observed that students taking the NW9 were audibly annoyed by having to complete the same measures for three times. Thus, it is possible that some students’ did not complete all measures honestly or appropriately. Since the data were collected from an operational testing environment, keeping the testing session authentic is a priority. Thus, to prevent students from irritation or distraction, other methods of collecting these data could be explored. Researchers have used a single item to measure students’ emotions (Goetz et al., 2007), which compared to using three items (as in the current study) would reduce the
interruption of taking the test. Additionally, students’ were required to complete measures on five emotions that I have decided to include in the current study. By repeatedly seeing items about each emotion, students may have been primed to feel a certain level of those emotions. The five emotions chosen in the current study were based on previous research. However, a better approach may have been to ask open-ended questions about which emotions the students are feeling at any given part of the test or have students select which emotions they are feeling from an exhaustive list of emotions. Future studies should explore different approaches to measuring students’ emotions and other constructs during an operational testing setting without inducing distraction.

**Conclusion**

The current study evaluated two theories regarding students’ motivation in low-stakes test context. The findings did not support the EV theory hypothesized effects between test-taking importance, examinee effort, and test performance. From a methodological standpoint, the present study offered advantages over many of the cross-sectional studies conducted to evaluate the EV theory hypotheses. The current study supported the CV theory hypotheses regarding the effects of test emotions on examinee effort and test performance. Future study should continue to test CV theory and its applicability in a low-stakes testing context.
Table 1

Three-dimensional model of achievement emotions

<table>
<thead>
<tr>
<th>Object of focus</th>
<th>Positive (Pleasant) Emotions</th>
<th>Negative (Unpleasant) Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Activating</td>
<td>Deactivating</td>
</tr>
<tr>
<td>Prospective Outcome</td>
<td>Hope, Anticipatory Joy</td>
<td>Relief</td>
</tr>
<tr>
<td></td>
<td>Joy</td>
<td></td>
</tr>
<tr>
<td>Retrospective Outcome</td>
<td>Pride, Gratitude</td>
<td>Relief, Contentment</td>
</tr>
<tr>
<td>Activity (Task)</td>
<td>Enjoyment</td>
<td>Relaxation</td>
</tr>
</tbody>
</table>

*Note.* Emotions are categorized by valence (positive vs. negative), physiological activation (activating vs. deactivating), and the object of focus (prospective outcome, retrospective outcome, and activity).
Table 2
Determinants and outcomes of achievement emotions

<table>
<thead>
<tr>
<th>Object Focus</th>
<th>Appraisal</th>
<th>Resulting Emotion</th>
<th>Effect on Cognitive-Motivational variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>High</td>
<td>Joy</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Anger</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Frustration</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Boredom/None</td>
<td>Negative</td>
</tr>
<tr>
<td>Prospective Outcome</td>
<td>High</td>
<td>Anticipatory Joy</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Hopelessness</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Relief</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>None</td>
<td>-</td>
</tr>
<tr>
<td>Retrospective Outcome</td>
<td>Self-success</td>
<td>High</td>
<td>Pride</td>
</tr>
<tr>
<td></td>
<td>Self-failure</td>
<td>High</td>
<td>Shame</td>
</tr>
<tr>
<td></td>
<td>Other-success</td>
<td>High</td>
<td>Gratitude</td>
</tr>
<tr>
<td></td>
<td>Other-failure</td>
<td>High</td>
<td>Anger</td>
</tr>
</tbody>
</table>

Note. According to CV theory, situational appraisals of perceived control over an achievement situation and the subjective value of an achievement situation will dictate what achievement emotions will develop (Pekrun, 2006). Rows differentiate the objects focus of each emotion, while columns display variations of levels of perceived control and the levels of subjective value that students may have. Taken together, different combinations of perceived control and subjective value will result in different emotions. Note that for Retrospective Outcome emotions, the person who exerts control, rather than the level of perceived control is important to consider. That is, control can be exerted by self (i.e., the student) or by others (i.e., teacher or peer). If the student experiences success and success is attributed to own actions, then pride will be experienced. If the student experiences failure and failure is attributed to own actions, then shame will be experienced. If success is experienced but other person’s action lead to it, gratitude will be experienced. If failure is experienced but other person’s actions lead to it, anger will be experienced. The last column refers to the effect of emotions on Cognitive-Motivational variables (e.g., motivation, learning strategies, etc. (see Figure 2)).
<table>
<thead>
<tr>
<th>Subjective Value</th>
<th>Expectancy-Value Theory</th>
<th>Control-Value Theory</th>
<th>Comparison of the different values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest</td>
<td>the intrinsic value that students experience as they complete an achievement task</td>
<td>Intrinsic - the importance of performing an activity for the sake of engaging in an achievement task</td>
<td>Similar</td>
</tr>
<tr>
<td>Utility (Usefulness)</td>
<td>the extrinsic value or utility that achievement activity can bring about</td>
<td>Extrinsic - the importance of an activity because attaining a favorable outcome is desirable</td>
<td>Similar</td>
</tr>
<tr>
<td>Attainment</td>
<td>personal relevance or gain that achievement task may facilitate for students</td>
<td>Achievement - the importance of doing well on a task for either intrinsic or extrinsic reasons</td>
<td>( Attainment ) value - important to achieve desirable outcome in order to preserve positive self-worth.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Achievement value - important to achieve desirable outcome for intrinsic and extrinsic reasons.</td>
<td></td>
</tr>
<tr>
<td>Research Design</td>
<td>Appropriate Time lags</td>
<td></td>
<td>Stationarity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------</td>
<td>-----------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td>Between direct effects</td>
<td>Between indirect effects</td>
<td>Stability</td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>Assumed</td>
<td>Assumed</td>
<td>Assumed</td>
</tr>
<tr>
<td>Sequential</td>
<td>Assumed</td>
<td>Assumed</td>
<td>Assumed</td>
</tr>
<tr>
<td>Half-Longitudinal</td>
<td>Yes</td>
<td>Assumed</td>
<td>Yes</td>
</tr>
<tr>
<td>Fully-Longitudinal</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fully-Longitudinal w/ 4+ waves of measurement</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Note.* The table provides information about certain concepts in mediation analysis. Cells that have “Assumed” indicate that concepts have to be assumed, meaning that there are no way to test them. Cells that have “Yes” in them indicate that concepts can be tested in particular research designs, thus they do not have to be assumed. Fully-Longitudinal research design with at least 4 waves of measurement allows to test all concepts discussed.
Table 5  
*Demographics of the sample*

<table>
<thead>
<tr>
<th></th>
<th>Assessment Day</th>
<th>Make-up Sessions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>38.29%</td>
<td>51.24%</td>
<td>41.57%</td>
</tr>
<tr>
<td>Female</td>
<td>61.71%</td>
<td>48.76%</td>
<td>58.43%</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>78.54%</td>
<td>75.21%</td>
<td>77.70%</td>
</tr>
<tr>
<td>Asian</td>
<td>7.01%</td>
<td>6.20%</td>
<td>6.61%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5.61%</td>
<td>9.09%</td>
<td>6.49%</td>
</tr>
<tr>
<td>Black</td>
<td>4.35%</td>
<td>7.02%</td>
<td>5.03%</td>
</tr>
<tr>
<td>American Indian</td>
<td>1.54%</td>
<td>0.83%</td>
<td>1.36%</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>0.70%</td>
<td>0.83%</td>
<td>0.73%</td>
</tr>
<tr>
<td>Not Specified</td>
<td>2.24%</td>
<td>0.83%</td>
<td>1.88%</td>
</tr>
<tr>
<td><strong>Mean age (years)</strong></td>
<td>18.45</td>
<td>18.5</td>
<td>18.46</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>713</td>
<td>242</td>
<td>955</td>
</tr>
</tbody>
</table>

*Note.* Students were able to select multiple ethnicities.
Table 6
Descriptive statistics and reliability for gathered variables

<table>
<thead>
<tr>
<th>Time 0</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Skewness</td>
<td>Kurtosis</td>
<td>Observed Min</td>
<td>Observed Max</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>14.034</td>
<td>2.719</td>
<td>-.084</td>
<td>.332</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>Value</td>
<td>16.835</td>
<td>3.084</td>
<td>-.276</td>
<td>.953</td>
<td>5</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>16.663</td>
<td>3.183</td>
<td>-.243</td>
<td>.696</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Effort</td>
<td>18.088</td>
<td>3.274</td>
<td>-.323</td>
<td>.151</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>Worry</td>
<td>7.936</td>
<td>2.723</td>
<td>.131</td>
<td>-.553</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Pride</td>
<td>8.735</td>
<td>1.989</td>
<td>-.186</td>
<td>.836</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Anger</td>
<td>7.334</td>
<td>2.464</td>
<td>.446</td>
<td>.125</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>6.529</td>
<td>2.200</td>
<td>.435</td>
<td>.340</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Boredom</td>
<td>10.096</td>
<td>2.664</td>
<td>-.180</td>
<td>-.218</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Subtest 1</td>
<td>14.115</td>
<td>3.183</td>
<td>-.165</td>
<td>-.274</td>
<td>5</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time 2</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>16.325</td>
<td>3.39</td>
<td>-.330</td>
<td>.763</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Effort</td>
<td>16.929</td>
<td>3.461</td>
<td>-.113</td>
<td>.233</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Worry</td>
<td>7.586</td>
<td>2.731</td>
<td>.269</td>
<td>-.305</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Pride</td>
<td>8.647</td>
<td>2.195</td>
<td>-.119</td>
<td>.732</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Anger</td>
<td>7.559</td>
<td>2.718</td>
<td>.415</td>
<td>-.009</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>6.407</td>
<td>2.275</td>
<td>.424</td>
<td>.137</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Boredom</td>
<td>10.16</td>
<td>2.857</td>
<td>-.254</td>
<td>-.294</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Subtest 2</td>
<td>15.006</td>
<td>3.782</td>
<td>-.633</td>
<td>.019</td>
<td>3</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time 3</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>16.254</td>
<td>3.466</td>
<td>-.297</td>
<td>.868</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Effort</td>
<td>16.724</td>
<td>3.568</td>
<td>-.077</td>
<td>.221</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Worry</td>
<td>7.527</td>
<td>2.684</td>
<td>.241</td>
<td>-.250</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Pride</td>
<td>8.622</td>
<td>2.346</td>
<td>-.230</td>
<td>.500</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>Effort</td>
<td>Self-Efficacy</td>
<td>Scale Range</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>--------</td>
<td>---------------</td>
<td>-----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>7.535</td>
<td>2.782</td>
<td>.437</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td>6.457</td>
<td>2.378</td>
<td>.417</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boredom</td>
<td>10.187</td>
<td>2.952</td>
<td>-.258</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtest 3</td>
<td>14.038</td>
<td>3.648</td>
<td>-.519</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The possible range of self-efficacy scores was from 3 to 21. The possible range of value and effort scores was from 5 to 25. The possible range of anger, boredom, enjoyment, pride, and worry was from 3 to 15. The possible range of subtest scores was from 0 to 22.
Table 7

*Proposed research questions, analyses, expected results, and the implications of results.*

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Analyses</th>
<th>Expected Results</th>
<th>Implications</th>
<th>Figures Depicting Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Do value, effort, emotions, performance relate to other constructs within or across time, after controlling for autoregressive effects?</td>
<td>Estimate Proposed Autoregressive Model A, which specifies all correlations are zero (other than the direct effects of a previous measure of a construct and the subsequent measure of a construct. Evaluate absolute model fit.</td>
<td>Proposed Autoregressive Model A will not fit the data well in an absolute sense.</td>
<td>Different constructs are related across and within-time</td>
<td>Figure 6</td>
</tr>
<tr>
<td>2. Do value, effort, emotions, performance relate within-time after controlling for autoregressive effects?</td>
<td>Estimate Proposed Autoregressive Model B, which only specifies autoregressive effects and correlated disturbances within-time. Evaluate the absolute model-data fit. And fit relative to proposed Autoregressive Model A</td>
<td>Proposed Autoregressive Models A or B will not fit the data well in an absolute sense. Proposed Autoregressive model B will fit the data better than Proposed Autoregressive model A</td>
<td>Constructs are related within-time, (i.e., significant disturbance correlations); thus, additional models can examine if previously measured constructs (e.g., anger at time 1) can explain subsequent within-time relations (e.g., correlation between anger and worry at time 2).</td>
<td>Figure 6, Figure 7</td>
</tr>
<tr>
<td>3. Do the effects of test-importance on examinee effort depend on self-efficacy?</td>
<td>Multiple regression analyses</td>
<td>Based on EV theory, the interaction terms should be significant.</td>
<td>The interaction between self-efficacy and test-importance should be modeled</td>
<td></td>
</tr>
<tr>
<td>4. Can EV theory explain relations between value, effort, emotions, and performance across time</td>
<td>Estimate Proposed EV theory Model A. Compare the fit of this model to the fit of the Proposed Autoregressive model B.</td>
<td>Proposed EV theory model A will fit the data better than the Proposed Autoregressive model B. Proposed EV theory model A will not fit the data well in an absolute sense</td>
<td>Other constructs affect examinee effort beyond self-efficacy and test-importance</td>
<td>Figure 7, Figure 8</td>
</tr>
<tr>
<td>5. Are the effects of self-efficacy on subsequent motivation important?</td>
<td>Estimate Proposed EV theory model B. Compare the fit to the fit of Proposed EV theory Model A.</td>
<td>Proposed EV theory model B should fit the data better than the Proposed EV theory model A*. However, neither models should fit well in an absolute sense.</td>
<td>While it is important to model the effects of self-efficacy on subsequent examinee effort, there are other constructs worth considering</td>
<td>Figure 8, Figure 9</td>
</tr>
</tbody>
</table>
6. Can the Combined EV and CV Theory Model explain the relations value, effort, emotions, and performance across time?

Estimate Proposed Combined EV & CV Model B. Evaluate absolute model-data fit.

Proposed Combined EV & CV Model B should fit the data poorly in an absolute sense.

Constructs other than self-efficacy, test-importance, and emotions affect subsequent examinee effort

Figure 11

7. Are the CV theory effects important when predicting examinee effort?

Compare the fit of Proposed Combined EV & CV theory model B to the fit of Proposed EV theory model B.

Proposed Combined EV & CV model B should fit the data better than the Proposed EV theory B.

Constructs specified in CV theory are important when explaining examinee effort, after controlling for EV theory

Figure 9, Figure 11

8. Are the EV theory effects important when predicting examinee effort?

Estimate Proposed CV theory model B. Compare the fit of this model to the fit of Proposed Combined EV & CV theory model B.

The fit of Proposed Combined EV & CV theory model B will not statistically differ from the Proposed CV theory model B.

Effects from EV theory are negligible when explaining examinee effort

Figure 10, Figure 12

9. Can the CV Theory Model explain the relations between value, effort, test emotions, and test performance across time?

Estimate the Proposed CV theory model B. Evaluate the absolute model-data fit.

The fit of Proposed CV theory model B will fit the data poorly in an absolute sense.

Constructs other than test-taking emotions may be important when explaining examinee effort

Figure 12

10. Can the Reciprocal Effects Models explain the relations between value, effort, test emotions, and test performance across time?

Estimate Proposed CV Adjacent reciprocal effects model B. Evaluate model-data fit, compare the fit to CV theory model B. Estimate Proposed CV complete reciprocal effects model B. Evaluate absolute model-data fit and compare it to Proposed CV Adjacent reciprocal effects model B.

The Proposed CV Adjacent reciprocal effects model B will fit the data better than Proposed CV theory model B. However, the model will not fit the data well in an absolute sense. The Proposed CV Complete reciprocal effects model B will fit the data better than the Proposed CV Adjacent reciprocal effects model B. Proposed CV Complete reciprocal effects model B will fit the data well in an absolute sense.

CV theory can be used to explain examinee effort. Further support for reciprocal effects between constructs specified in CV theory.

Figure 12, Figure 14, Figure 16

Note. * It is expected that Proposed EV theory model B will fit the data better than Proposed EV Theory model A. However, if the fit of both models is statistically equivalent, then models without the effects of self-efficacy at Time 0 on constructs at Time 2 and Time 3 will be used. These models are termed "A", answer the same research questions, and are depicted in Figures 9, 11, 13, and 15.
Table 8

*Actual research questions, analyses, expected results, and the implications of results.*

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Analyses</th>
<th>Expected Results</th>
<th>Implications</th>
<th>Figures Depicting Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Can value, effort, emotion, and performance relations across time be adequately explained by autoregressive effects</td>
<td>Estimate Autoregressive Model A, which specifies all correlations are zero, other than the within-time correlation at time 1 and across-time correlations between the same construct over time. Evaluate absolute model fit.</td>
<td>Autoregressive Model A will not fit the data well.</td>
<td>Different constructs are related across and within-time</td>
<td>Figure 18</td>
</tr>
<tr>
<td>2. Can value, effort, emotion, and performance relations across time be adequately explained via autoregressive effects and within-time correlations?</td>
<td>Estimate Autoregressive Model B, which only specifies autoregressive effects and correlations between constructs within-time. Evaluate the absolute model-data fit. And fit relative to Autoregressive Model A.</td>
<td>Autoregressive Models B will not fit the data well in an absolute sense. Autoregressive Model B will fit the data better than Autoregressive model A.</td>
<td>Constructs are related within-time. (i.e., significant disturbance correlations); thus, subsequent models can examine if previously measured constructs (e.g., anger at time 1) can explain subsequent within-time relations (e.g., correlation between anger and worry at time 2).</td>
<td>Figure 18, Figure 19</td>
</tr>
<tr>
<td>3. Can EV theory explain relations between value, effort, emotions, and performance across time?</td>
<td>Estimate EV theory Model. Compare the fit of this model to the fit of the Autoregressive model B.</td>
<td>EV Theory Model will fit the data better than Autoregressive Model B. EV Theory Model will not fit the data well in an absolute sense.</td>
<td>Other constructs (e.g., emotions) affect examinee effort beyond perceived value.</td>
<td>Figure 19, Figure 20</td>
</tr>
<tr>
<td>4. Can the Combined EV and CV Theory Model explain the relations between value, effort, emotions, and performance across time</td>
<td>Estimate Combined EV and CV Model. Compare the model-data fit to EV Theory Model and evaluate absolute model-data fit.</td>
<td>Combined EV and CV Model will fit the data poorly in an absolute sense but it will fit better than the EV Theory Model</td>
<td>Constructs specified in CV theory are important when predicting examinee effort, after controlling for EV theory.</td>
<td>Figure 21</td>
</tr>
<tr>
<td></td>
<td>Question</td>
<td>Methodology</td>
<td>Evaluation</td>
<td>Figure</td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>5.</td>
<td>Are the EV theory effects important when predicting examinee effort?</td>
<td>Compare the fit of Combined EV and CV Theory Model to the fit of CV Theory Model. Evaluate</td>
<td>The fit of Combined EV &amp; CV Theory Model will not statistically differ from the CV Theory Model.</td>
<td>Figure 21, Figure 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Constructs specified in EV theory are not important when predicting examinee effort</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Can the CV Theory Model explain the relations between value, effort, test emotions, and test performance across time?</td>
<td>Estimate CV Theory Model. Evaluate absolute model-data fit.</td>
<td>CV Theory Model will fit the data poorly.</td>
<td>Figure 22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Basic CV Theory cannot explain relations between construct across time.</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Can the Reciprocal Effects Models explain the relations between value, effort, test emotions, and test performance across time?</td>
<td>Estimate Adjacent Reciprocal Effects Model and evaluate model-data fit. Compare the fit of Adjacent Reciprocal Effects Model to CV Theory Model. Estimate Complete Reciprocal Effects Model and evaluate model-data fit. Compare Complete Reciprocal Effects Model to Adjacent Reciprocal Effects Model.</td>
<td>C The Adjacent Reciprocal Effects Model will fit the data better than CV Theory Model. However, the model will not fit the data well in an absolute sense. The Complete Reciprocal Effects Model will fit the data better than the Adjacent Reciprocal Effects Model. Complete Reciprocal Effects Model will fit the data well in an absolute sense.</td>
<td>Figure 22, Figure 23, Figure 24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Relations between constructs over time can be explained by reciprocal effects, as specified in CV Theory.</td>
<td></td>
</tr>
</tbody>
</table>

*Note. The table follows the same structure as Table 7. However, due to inability to model self-efficacy, several research questions could not be tested.*
### Model summary for mean stability

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>464.422</td>
<td>16</td>
<td>0.998</td>
<td>0.149</td>
</tr>
<tr>
<td>Effort</td>
<td>295.424</td>
<td>2</td>
<td>0.986</td>
<td>0.392</td>
</tr>
<tr>
<td>Test Performance</td>
<td>101.68</td>
<td>2</td>
<td>0.995</td>
<td>0.228</td>
</tr>
<tr>
<td>Test Emotions</td>
<td>87.639</td>
<td>10</td>
<td>0.996</td>
<td>0.0902</td>
</tr>
<tr>
<td>Anger</td>
<td>19.323</td>
<td>2</td>
<td>0.999</td>
<td>0.0952</td>
</tr>
<tr>
<td>Boredom</td>
<td>1.941*</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>6.448</td>
<td>2</td>
<td>1</td>
<td>0.0483</td>
</tr>
<tr>
<td>Pride</td>
<td>6.206</td>
<td>2</td>
<td>1</td>
<td>0.0469</td>
</tr>
<tr>
<td>Worry</td>
<td>54.594</td>
<td>2</td>
<td>0.997</td>
<td>0.166</td>
</tr>
</tbody>
</table>

*Note.* *statistically significant at .05.*
Table 10

*Bivariate correlations between the same constructs at different times*

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Importance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.853</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.812</td>
<td>0.866</td>
<td>1</td>
</tr>
<tr>
<td><strong>Effort</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.805</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.755</td>
<td>0.879</td>
<td>1</td>
</tr>
<tr>
<td><strong>Worry</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.831</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.778</td>
<td>0.882</td>
<td>1</td>
</tr>
<tr>
<td><strong>Pride</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.818</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.777</td>
<td>0.862</td>
<td>1</td>
</tr>
<tr>
<td><strong>Anger</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.811</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.779</td>
<td>0.865</td>
<td>1</td>
</tr>
<tr>
<td><strong>Enjoy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.777</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.735</td>
<td>0.840</td>
<td>1</td>
</tr>
<tr>
<td><strong>Boredom</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.802</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.734</td>
<td>0.852</td>
<td>1</td>
</tr>
<tr>
<td><strong>Test Performance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>0.546</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>0.528</td>
<td>0.639</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 11  
*Model Summary for evaluating equality of variances*

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>21.825</td>
<td>2</td>
<td>0.999</td>
<td>0.102</td>
<td>0.013</td>
</tr>
<tr>
<td>Effort</td>
<td>16.253</td>
<td>2</td>
<td>0.999</td>
<td>0.086</td>
<td>0.013</td>
</tr>
<tr>
<td>Test Performance</td>
<td>41.996</td>
<td>2</td>
<td>0.998</td>
<td>0.145</td>
<td>0.019</td>
</tr>
<tr>
<td>Test Emotions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>37.801</td>
<td>2</td>
<td>0.998</td>
<td>0.137</td>
<td>0.018</td>
</tr>
<tr>
<td>Boredom</td>
<td>21.999</td>
<td>2</td>
<td>0.999</td>
<td>0.102</td>
<td>0.014</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>13.302</td>
<td>2</td>
<td>0.999</td>
<td>0.077</td>
<td>0.010</td>
</tr>
<tr>
<td>Pride</td>
<td>63.823</td>
<td>2</td>
<td>0.997</td>
<td>0.180</td>
<td>0.022</td>
</tr>
<tr>
<td>Worry</td>
<td>1.322*</td>
<td>2</td>
<td>1.000</td>
<td>0.000</td>
<td>0.002</td>
</tr>
</tbody>
</table>

*Note.* *statistically significant at .05
<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>Correlation residuals &gt;</th>
<th>Correlation residuals &gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoregressive model A</td>
<td>2186.9</td>
<td>232</td>
<td>0.907</td>
<td>0.0939</td>
<td>0.1160</td>
<td>74</td>
<td>39</td>
</tr>
<tr>
<td>Autoregressive model B</td>
<td>1151.17</td>
<td>176</td>
<td>0.954</td>
<td>0.0762</td>
<td>0.1150</td>
<td>98</td>
<td>24</td>
</tr>
<tr>
<td>EV Theory Model</td>
<td>1067.41</td>
<td>172</td>
<td>0.957</td>
<td>0.0738</td>
<td>0.1010</td>
<td>86</td>
<td>13</td>
</tr>
<tr>
<td>Combined EV and CV Theory model</td>
<td>912.624</td>
<td>152</td>
<td>0.964</td>
<td>0.0724</td>
<td>0.0694</td>
<td>35</td>
<td>3</td>
</tr>
<tr>
<td>CV Theory Model</td>
<td>929.304</td>
<td>154</td>
<td>0.963</td>
<td>0.0726</td>
<td>0.0744</td>
<td>41</td>
<td>4</td>
</tr>
<tr>
<td>Adjacent Reciprocal Effects Model</td>
<td>770.09</td>
<td>132</td>
<td>0.97</td>
<td>0.0711</td>
<td>0.0490</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Complete Reciprocal Effects Model</td>
<td>700.144</td>
<td>104</td>
<td>0.972</td>
<td>0.0775</td>
<td>0.0438</td>
<td>15</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 13
*Direct effects of constructs at time 1 on constructs at time 2 from Adjacent Reciprocal Effects Model*

<table>
<thead>
<tr>
<th>Predictor at Time 1</th>
<th>Outcome at Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perceived Value</td>
</tr>
<tr>
<td>Perceived Value</td>
<td>0.837*</td>
</tr>
<tr>
<td></td>
<td>0.020</td>
</tr>
<tr>
<td>Anger</td>
<td>0.100*</td>
</tr>
<tr>
<td></td>
<td>0.027</td>
</tr>
<tr>
<td>Boredom</td>
<td>0.059*</td>
</tr>
<tr>
<td></td>
<td>0.026</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>0.029</td>
</tr>
<tr>
<td>Pride</td>
<td>0.093*</td>
</tr>
<tr>
<td></td>
<td>0.034</td>
</tr>
<tr>
<td>Worry</td>
<td>0.096*</td>
</tr>
<tr>
<td></td>
<td>0.023</td>
</tr>
<tr>
<td>Effort</td>
<td>-0.074*</td>
</tr>
<tr>
<td></td>
<td>0.018</td>
</tr>
<tr>
<td>Test Performance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.018</td>
</tr>
</tbody>
</table>

Note. * statistically significant at .05
Table 14

Direct effects of constructs at time 2 on constructs at time 3 from Adjacent Reciprocal Effects Model

<table>
<thead>
<tr>
<th>Predictor at Time 2</th>
<th>Perceived Value</th>
<th>Anger</th>
<th>Boredom</th>
<th>Enjoyment</th>
<th>Pride</th>
<th>Worry</th>
<th>Effort</th>
<th>Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Value</td>
<td>Unstandardized effect</td>
<td>0.828*</td>
<td>-0.051*</td>
<td>-0.042*</td>
<td>0.01</td>
<td>0.01</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>0.019</td>
<td>0.015</td>
<td>0.017</td>
<td>0.014</td>
<td>0.013</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>Unstandardized effect</td>
<td>-0.021</td>
<td>0.819*</td>
<td></td>
<td></td>
<td></td>
<td>-0.019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>0.025</td>
<td>0.019</td>
<td></td>
<td></td>
<td></td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td>Boredom</td>
<td>Unstandardized effect</td>
<td>-0.068*</td>
<td>0.847*</td>
<td></td>
<td></td>
<td></td>
<td>0.117*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>0.024</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td>Unstandardized effect</td>
<td>0.013</td>
<td></td>
<td>0.845*</td>
<td></td>
<td></td>
<td>-0.049</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>0.028</td>
<td>0.019</td>
<td></td>
<td></td>
<td></td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>Pride</td>
<td>Unstandardized effect</td>
<td>0.100*</td>
<td></td>
<td></td>
<td>0.864*</td>
<td></td>
<td>0.143*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>0.029</td>
<td></td>
<td></td>
<td>0.019</td>
<td></td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Worry</td>
<td>Unstandardized effect</td>
<td>0.078*</td>
<td></td>
<td></td>
<td>0.846*</td>
<td></td>
<td>0.074*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>0.022</td>
<td></td>
<td></td>
<td>0.015</td>
<td></td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>Effort</td>
<td>Unstandardized effect</td>
<td>-0.033*</td>
<td>-0.006</td>
<td>0.019</td>
<td>0.052*</td>
<td>-0.023</td>
<td>0.802*</td>
<td>0.157*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>0.015</td>
<td>0.018</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.021</td>
<td>0.027</td>
</tr>
<tr>
<td>Test Performance</td>
<td>Unstandardized effect</td>
<td></td>
<td></td>
<td></td>
<td>0.038*</td>
<td></td>
<td>0.565*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td></td>
<td></td>
<td></td>
<td>0.015</td>
<td></td>
<td>0.025</td>
<td></td>
</tr>
</tbody>
</table>

Note. * statistically significant at .05
Table 15

*Indirect effects of constructs at time 1 on constructs at time 3 from Adjacent Reciprocal Effects Model*

<table>
<thead>
<tr>
<th>Predictor at Time 1</th>
<th>Outcome at Time 3</th>
<th>Perceived Value</th>
<th>Anger</th>
<th>Boredom</th>
<th>Enjoyment</th>
<th>Pride</th>
<th>Worry</th>
<th>Effort</th>
<th>Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Value</td>
<td></td>
<td>Unstandardized effect</td>
<td>.699*</td>
<td>-.083*</td>
<td>-.028</td>
<td>.039*</td>
<td>.027</td>
<td>.024</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>.023</td>
<td>.020</td>
<td>.021</td>
<td>.018</td>
<td>.016</td>
<td>.019</td>
<td>.004</td>
</tr>
<tr>
<td>Anger</td>
<td></td>
<td>Unstandardized effect</td>
<td>-.100*</td>
<td>.670*</td>
<td>.005*</td>
<td>-.002</td>
<td>-.005*</td>
<td>.001</td>
<td>-.071*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>.030</td>
<td>.023</td>
<td>.002</td>
<td>.001</td>
<td>.002</td>
<td>.001</td>
<td>.030</td>
</tr>
<tr>
<td>Boredom</td>
<td></td>
<td>Unstandardized effect</td>
<td>-.103*</td>
<td>.008*</td>
<td>.684*</td>
<td>-.004</td>
<td>-.009*</td>
<td>.003</td>
<td>-.221*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>.028</td>
<td>.003</td>
<td>.023</td>
<td>.002</td>
<td>.003</td>
<td>.002</td>
<td>.030</td>
</tr>
<tr>
<td>Enjoyment</td>
<td></td>
<td>Unstandardized effect</td>
<td>.012</td>
<td>.001</td>
<td>.000</td>
<td>.638*</td>
<td>-.002</td>
<td>.001</td>
<td>-.068*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>.032</td>
<td>.002</td>
<td>.001</td>
<td>.022</td>
<td>.002</td>
<td>.001</td>
<td>.033</td>
</tr>
<tr>
<td>Pride</td>
<td></td>
<td>Unstandardized effect</td>
<td>.160*</td>
<td>-.010*</td>
<td>-.005</td>
<td>.004</td>
<td>.727*</td>
<td>-.004</td>
<td>.253*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>.037</td>
<td>.003</td>
<td>.003</td>
<td>.002</td>
<td>.024</td>
<td>.002</td>
<td>.039</td>
</tr>
<tr>
<td>Worry</td>
<td></td>
<td>Unstandardized effect</td>
<td>.143*</td>
<td>-.008*</td>
<td>-.004*</td>
<td>.002</td>
<td>.005*</td>
<td>.682*</td>
<td>.123*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>.026</td>
<td>.002</td>
<td>.002</td>
<td>.001</td>
<td>.002</td>
<td>.020</td>
<td>.026</td>
</tr>
<tr>
<td>Effort</td>
<td></td>
<td>Unstandardized effect</td>
<td>.007</td>
<td>-.085*</td>
<td>-.057*</td>
<td>.002</td>
<td>.092*</td>
<td>-.068*</td>
<td>.608*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>.004</td>
<td>.018</td>
<td>.021</td>
<td>.017</td>
<td>.016</td>
<td>.018</td>
<td>.024</td>
</tr>
<tr>
<td>Test Performance</td>
<td></td>
<td>Unstandardized effect</td>
<td>-.002</td>
<td>.000</td>
<td>.001</td>
<td>.003*</td>
<td>-.001</td>
<td>.063*</td>
<td>.344*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard Error</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.001</td>
<td>.018</td>
<td>.023</td>
</tr>
</tbody>
</table>

*Note.* *statistically significant at .05
Table 16

*Direct effects of constructs at time 1 on constructs at time 2 from Complete Reciprocal Effects Model*

<table>
<thead>
<tr>
<th></th>
<th>Perceived Value</th>
<th>Anger</th>
<th>Boredom</th>
<th>Enjoyment</th>
<th>Pride</th>
<th>Worry</th>
<th>Effort</th>
<th>Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceived Value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstandardized effect</td>
<td>.831*</td>
<td>-.052*</td>
<td>.001</td>
<td>.040*</td>
<td>.029*</td>
<td>.026</td>
<td>.031</td>
<td>.069</td>
</tr>
<tr>
<td>Standard Error</td>
<td>.021</td>
<td>.019</td>
<td>.020</td>
<td>.017</td>
<td>.015</td>
<td>.018</td>
<td>.024</td>
<td>.038</td>
</tr>
<tr>
<td><strong>Anger</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstandardized effect</td>
<td>-.092*</td>
<td>.808*</td>
<td></td>
<td></td>
<td></td>
<td>-.067*</td>
<td>-.121*</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>.028</td>
<td>.022</td>
<td></td>
<td></td>
<td></td>
<td>.030</td>
<td>.050</td>
<td></td>
</tr>
<tr>
<td><strong>Boredom</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstandardized effect</td>
<td>-.051</td>
<td>.800*</td>
<td></td>
<td></td>
<td></td>
<td>-.154*</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>.026</td>
<td>.021</td>
<td></td>
<td></td>
<td></td>
<td>.030</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td><strong>Enjoyment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstandardized effect</td>
<td>.008</td>
<td></td>
<td>.756*</td>
<td></td>
<td>-.044</td>
<td>-.086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>.030</td>
<td></td>
<td>.020</td>
<td></td>
<td>.033</td>
<td>.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pride</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstandardized effect</td>
<td>.075*</td>
<td></td>
<td></td>
<td>.829*</td>
<td>.159*</td>
<td>.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>.035</td>
<td></td>
<td></td>
<td>.022</td>
<td>.039</td>
<td>.064</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Worry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstandardized effect</td>
<td>.100*</td>
<td></td>
<td></td>
<td></td>
<td>.799*</td>
<td>.073*</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>.023</td>
<td></td>
<td></td>
<td></td>
<td>.019</td>
<td>.026</td>
<td>.042</td>
<td></td>
</tr>
<tr>
<td><strong>Effort</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstandardized effect</td>
<td>.031</td>
<td>-.075*</td>
<td>-.066*</td>
<td>-.010</td>
<td>.065*</td>
<td>-.053*</td>
<td>.727*</td>
<td>.194*</td>
</tr>
<tr>
<td>Standard Error</td>
<td>.022</td>
<td>.019</td>
<td>.021</td>
<td>.017</td>
<td>.015</td>
<td>.018</td>
<td>.025</td>
<td>.040</td>
</tr>
<tr>
<td><strong>Test Performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstandardized effect</td>
<td>.017</td>
<td>-.013</td>
<td>.010</td>
<td>-.013</td>
<td>-.017</td>
<td>-.042*</td>
<td>.048*</td>
<td>.584*</td>
</tr>
<tr>
<td>Standard Error</td>
<td>.019</td>
<td>.016</td>
<td>.018</td>
<td>.015</td>
<td>.013</td>
<td>.016</td>
<td>.021</td>
<td>.033</td>
</tr>
</tbody>
</table>

*Note. * statistically significant at .05
### Table 17

**Direct effects of constructs at time 2 on constructs at time 3 from Complete Reciprocal Effects Model**

<table>
<thead>
<tr>
<th>Predictor at Time 2</th>
<th>Perceived Value</th>
<th>Anger</th>
<th>Boredom</th>
<th>Enjoyment</th>
<th>Pride</th>
<th>Worry</th>
<th>Effort</th>
<th>Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Value</td>
<td>Unstandardized effect</td>
<td>.813*</td>
<td>-.053*</td>
<td>-.044*</td>
<td>.011</td>
<td>.012</td>
<td>.002</td>
<td>.037</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.020</td>
<td>.016</td>
<td>.017</td>
<td>.015</td>
<td>.013</td>
<td>.014</td>
<td>.020</td>
</tr>
<tr>
<td>Anger</td>
<td>Unstandardized effect</td>
<td>.001</td>
<td>.815*</td>
<td></td>
<td></td>
<td></td>
<td>-.001</td>
<td>-.002</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.025</td>
<td>.019</td>
<td></td>
<td></td>
<td></td>
<td>.024</td>
<td>.041</td>
</tr>
<tr>
<td>Boredom</td>
<td>Unstandardized effect</td>
<td>-.028</td>
<td>.842*</td>
<td>-.101*</td>
<td>.103*</td>
<td>.024</td>
<td>.041</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.025</td>
<td>.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enjoyment</td>
<td>Unstandardized effect</td>
<td>.033</td>
<td></td>
<td>.844*</td>
<td></td>
<td>-.048</td>
<td>.024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.028</td>
<td>.019</td>
<td></td>
<td></td>
<td>.027</td>
<td>.046</td>
<td></td>
</tr>
<tr>
<td>Pride</td>
<td>Unstandardized effect</td>
<td>.045</td>
<td></td>
<td></td>
<td>.860*</td>
<td>.119*</td>
<td>.050*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.031</td>
<td>.020</td>
<td></td>
<td></td>
<td>.030</td>
<td>.051</td>
<td></td>
</tr>
<tr>
<td>Worry</td>
<td>Unstandardized effect</td>
<td>.069*</td>
<td></td>
<td></td>
<td>.849*</td>
<td>.053*</td>
<td>-.123*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.022</td>
<td>.015</td>
<td></td>
<td></td>
<td>.021</td>
<td>.036</td>
<td></td>
</tr>
<tr>
<td>Effort</td>
<td>Unstandardized effect</td>
<td>.089*</td>
<td>-.039*</td>
<td>-.014</td>
<td>.026</td>
<td>.059*</td>
<td>-.025</td>
<td>.813*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.022</td>
<td>.016</td>
<td>.019</td>
<td>.015</td>
<td>.014</td>
<td>.014</td>
<td>.021</td>
</tr>
<tr>
<td>Test Performance</td>
<td>Unstandardized effect</td>
<td>.009</td>
<td>.001</td>
<td>.003</td>
<td>-.016</td>
<td>-.007</td>
<td>.005</td>
<td>.036*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.016</td>
<td>.013</td>
<td>.014</td>
<td>.012</td>
<td>.011</td>
<td>.012</td>
<td>.015</td>
</tr>
</tbody>
</table>

*Note.* *statistically significant at .05
Table 18

*Indirect effects of constructs at time 1 on construct at time 3 from Complete Reciprocal Effects Model*

<table>
<thead>
<tr>
<th>Predictor at Time 1</th>
<th>Perceived Value</th>
<th>Anger</th>
<th>Boredom</th>
<th>Enjoyment</th>
<th>Pride</th>
<th>Worry</th>
<th>Effort</th>
<th>Test Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Value</td>
<td>Unstandardized effect</td>
<td>.683*</td>
<td>.088*</td>
<td>-.036</td>
<td>.042*</td>
<td>.036*</td>
<td>.023</td>
<td>.061*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>-.024</td>
<td>.020</td>
<td>.022</td>
<td>.018</td>
<td>.017</td>
<td>.019</td>
<td>.026</td>
</tr>
<tr>
<td>Anger</td>
<td>Unstandardized effect</td>
<td>-.081*</td>
<td>.666*</td>
<td>.005</td>
<td>.001</td>
<td>-.004</td>
<td>.001</td>
<td>-.063*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.030</td>
<td>.023</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td>.031</td>
</tr>
<tr>
<td>Boredom</td>
<td>Unstandardized effect</td>
<td>-.077*</td>
<td>.009*</td>
<td>.679*</td>
<td>-.005*</td>
<td>-.010*</td>
<td>.004</td>
<td>-.206*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.029</td>
<td>.003</td>
<td>.023</td>
<td>.003</td>
<td>.003</td>
<td>.002</td>
<td>.031</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>Unstandardized effect</td>
<td>.027</td>
<td>.001</td>
<td>.000</td>
<td>.638*</td>
<td>-.002</td>
<td>.001</td>
<td>-.075*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.032</td>
<td>.002</td>
<td>.002</td>
<td>.022</td>
<td>.002</td>
<td>.001</td>
<td>.034</td>
</tr>
<tr>
<td>Pride</td>
<td>Unstandardized effect</td>
<td>.112*</td>
<td>-.010*</td>
<td>-.006</td>
<td>.005</td>
<td>.724*</td>
<td>.004</td>
<td>.231*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.038</td>
<td>.003</td>
<td>.003</td>
<td>.003</td>
<td>.025</td>
<td>.002</td>
<td>.040</td>
</tr>
<tr>
<td>Worry</td>
<td>Unstandardized effect</td>
<td>.143*</td>
<td>-.008*</td>
<td>-.005*</td>
<td>.003</td>
<td>.006*</td>
<td>.676*</td>
<td>.106*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.026</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td>.002</td>
<td>.020</td>
<td>.027</td>
</tr>
<tr>
<td>Effort</td>
<td>Unstandardized effect</td>
<td>.093*</td>
<td>-.090*</td>
<td>-.067*</td>
<td>.008</td>
<td>.098*</td>
<td>-.062*</td>
<td>.611*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.024</td>
<td>.019</td>
<td>.021</td>
<td>.017</td>
<td>.016</td>
<td>.018</td>
<td>.026</td>
</tr>
<tr>
<td>Test Performance</td>
<td>Unstandardized effect</td>
<td>.019</td>
<td>-.012</td>
<td>.009</td>
<td>-.018</td>
<td>-.016</td>
<td>-.034*</td>
<td>.055*</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>.018</td>
<td>.016</td>
<td>.017</td>
<td>.014</td>
<td>.013</td>
<td>.015</td>
<td>.020</td>
</tr>
</tbody>
</table>

*Note. * statistically significant at .05*
Table 19

Summary of changes for the direct effects between the two reciprocal effects models

<table>
<thead>
<tr>
<th>The Effect of…</th>
<th>…on</th>
<th>Significant in the Adjacent Reciprocal Effects Model</th>
<th>Significant in the Complete Reciprocal Effects Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boredom at time 1</td>
<td>Perceived Value at time 2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Boredom at time 2</td>
<td>Perceived Value at time 3</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Boredom at time 1</td>
<td>Test Performance at time 3</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Anger at time 1</td>
<td>Boredom at time 3</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Anger at time 1</td>
<td>Pride at time 3</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Anger at time 1</td>
<td>Test Performance at time 3</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Pride at time 1</td>
<td>Test Performance at time 3</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Perceived value at time 1</td>
<td>Pride at time 2</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Perceived value at time 1</td>
<td>Pride at time 3</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Perceived value at time 1</td>
<td>Effort at time 3</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Boredom at time 1</td>
<td>Enjoyment at time 3</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Effort at time 1</td>
<td>Perceived value at time 3</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Test performance at time 1</td>
<td>Worry at time 3</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Effort at time 2</td>
<td>Perceived value at time 3</td>
<td>Not estimated</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Figure 1. The Expectancy-Value (EV) theory of examinee effort. According to EV theory, students’ expectancy to succeed, intrinsic value, attainment value, utility values, and cost influences students’ motivation, which then affects task performance.
Figure 2. The Control-Value (CV) theory of achievement emotions. According to CV theory, students’ environment causes situational appraisals, which then cause achievement emotions. Achievement emotions are hypothesized to exert an effect on achievement performance via intervening cognitive-motivational variables. Reciprocal effects (indicated by the multi-headed arrow at the top of the figure) exist in the CV theory. For example, situational appraisals are thought to influence subsequent emotions, however, emotions are then hypothesized to shape subsequent situational appraisals. Reciprocal effects are present throughout the whole model.
Figure 3. Depiction of simple regression (a), partial mediation (b), and complete mediation (c). Each model exemplifies causal effects of X on Y, however the theory underlying the effects is different in each case. In model A, the theory is that X causes Y directly. In model B the theory specifies that X is cause Y by exerting its influence via direct effect (c’) and indirectly via M (a*b). In model C, the theory is that X causes Y only indirectly via M (a*b).
Figure 4. Popular research designs for estimating mediation: A) Cross-sectional, B) Sequential, C) Half-Longitudinal, D) Fully-Longitudinal. Notice that each design differs in the number of measurement occasions.
Figure 5. A path model with five waves of measurement. The figure depicts two different types of effects. First, consider the effects of importance at time 1 to importance at time 2 ($II_{12}$). This effect is called autoregressive effect, which depicts how the same construct has changed over time. Notice that there are many autoregressive effects in this model. To ease understanding, all autoregressive effects are depicted as the same first letter of each construct, when the numbers refer to the time of measurement. The second type of effect is called cross-lagged, which depicts an effect of one construct at time $t-1$ on a different construct at time $t$. For example, the effect of Anger at time 1 on Effort at time 2 is depicted as the path $AE_{12}$. The marking rationale used for cross-lagged effects is the same for autoregressive effects.
Figure 6. Proposed Autoregressive Model A. There are 307 degrees of freedom in this model. Both this and the subsequent, Proposed Autoregressive model B will serve as baseline for the subsequent model comparisons in which different hypotheses will be tested.
Figure 7. Proposed Autoregressive Model B. This model builds on the Proposed Autoregressive model A by adding 84 correlations between disturbance terms of all endogenous variables resulting in 223 degrees of freedom. Both Proposed Autoregressive model A and this model will serve as baseline for the subsequent model comparisons in which different hypotheses will be tested.
Figure 8. Proposed EV Theory Model A. This model builds on the Proposed Autoregressive model B by estimating the effects of value and self-efficacy on subsequent motivation. Additionally, the effects of motivation on subsequent performance are estimated. This model has 217 degrees of freedom.
Figure 9. Proposed EV Theory Model B. This model builds on the Proposed EV theory model A by estimating the effects of self-efficacy at time 0 (depicted in blue) on motivation at time 2 and time 3. This model has 215 degrees of freedom.
Figure 10. Proposed Combined EV and CV Theory Model A. This model builds on the Proposed EV theory model A by estimating the effects of value and self-efficacy on subsequent test emotions and by estimating the effects of test emotions on subsequent motivation. This model has 187 degrees of freedom.
Figure 11. Proposed Combined EV and CV Theory Model B. This model builds on the Proposed Combined EV and CV model A by estimating the effects of self-efficacy at time 0 (depicted in blue) on motivation at time 2 and time 3. This model has 185 degrees of freedom.
Figure 12. Proposed CV Theory Model A. Notice, Proposed CV model A is simpler than the Proposed Combined EV and CV model A. The Proposed EV theory effects (i.e., the effect of value and self-efficacy on subsequent motivation) are set to 0. This model builds on the Proposed Combined EV and CV model A by estimating the effects of self-efficacy on subsequent motivation. This model has 191 degrees of freedom.
Figure 13. Proposed CV Theory Model B. Notice, Proposed CV model B is simpler than the Proposed Combined EV and CV model B. The EV theory effects (i.e., the effect of value and self-efficacy on subsequent motivation) are set to 0. However, building on Proposed CV model A, the effects of self-efficacy at time 0 on test emotions at time 2 and time 3 (depicted in blue) are estimated. This model has 181 degrees of freedom.
Figure 14. Proposed Adjacent Reciprocal Effects Model A. This model builds on the Proposed CV theory model A by estimating reciprocal effects of “adjacent” constructs in the CV causal chain. That is, the effects of test emotions on subsequent value, the effects of motivation on subsequent emotions, and the effects of performance on subsequent motivation are estimated. All “adjacent” reciprocal effects are depicted in orange. This model has 169 degrees of freedom.
Figure 15. Proposed Adjacent Reciprocal Effects Model B. This model builds on the Proposed CV Adjacent reciprocal effects model A by estimating the effects of self-efficacy on subsequent test emotions (depicted in blue). This model has 159 degrees of freedom.
Figure 16. Proposed Complete Reciprocal Effects Model A. This model builds on the Proposed CV Adjacent reciprocal effects model A by estimating the following effects (depicted in green): the effects perceived value and self-efficacy on subsequent motivation and on subsequent performance, the effects of test emotions on subsequent performance, the effects of motivation on subsequent value, the effects of test performance on subsequent test emotions and subsequent value. This model has 137 degrees of freedom.
Figure 17. Proposed Complete Reciprocal Effects Model B. This model builds on the Proposed CV Complete reciprocal effects model B by estimating the effects of self-efficacy at time 0 (depicted in blue) on test emotions at time 2 and time 3. This model has 127 degrees of freedom.
Figure 18. Autoregressive model A. There are 232 degrees of freedom in this model. Note, the constructs at time 1 were allowed to correlate. Both this and the subsequent, Autoregressive model B will serve as baseline for the subsequent model comparisons in which different hypotheses will be tested.
Figure 19. Autoregressive Model B. This model builds on the Autoregressive model A by adding 56 correlations between disturbance terms of all endogenous variables resulting in 176 degrees of freedom. Both Autoregressive model A and this model will serve as baseline for the subsequent model comparisons in which different hypotheses will be tested.
Figure 20. EV Theory Model. This model builds on the Autoregressive model B by estimating the effects of value on subsequent motivation. Additionally, the effects of motivation on subsequent performance are estimated. This model has 172 degrees of freedom.
Figure 21. Combined EV and CV Theory Model. This model builds on the EV Theory Model by estimating the effects of value on subsequent test emotions and by estimating the effects of test emotions on subsequent motivation. This model has 152 degrees of freedom.
Figure 22. CV Theory Model. Notice, CV Theory Model is simpler than the Combined EV and CV Theory Model. The EV theory effects (i.e., the effect of value on subsequent motivation) are set to 0. This model has 154 degrees of freedom.
Figure 23. Adjacent Reciprocal Effects Model. This model builds on the CV Theory Model by estimating reciprocal effects of “adjacent” constructs in the CV causal chain. That is, the effects of test emotions on subsequent value, the effects of motivation on subsequent emotions, and the effects of performance on subsequent motivation are estimated. All “adjacent” reciprocal effects are depicted in orange. This model has 132 degrees of freedom.
Figure 24. Complete Reciprocal Effects Model. This model builds on the Adjacent Reciprocal Effects Model by estimating the following effects (depicted in green): the effects perceived value on subsequent motivation and on subsequent performance, the effects of test emotions on subsequent performance, the effects of motivation on subsequent value, the effects of test performance on subsequent test emotions and subsequent value. This model has 104 degrees of freedom.
Appendix A

Students saw the following test instructions before they began NW9. Bolded words at the end of each item were added in this appendix to indicate what construct each item represents. * indicates reverse coded items. Students did not see this.

Before we begin, here is an example item from the Natural World Test.

Example Item:

Goldstar Inc. claims that its SAT preparation course is superior to the course offered by Premiere Inc. A study conducted by Goldstar compared SAT scores from 500 students who took Goldstar’s course and 500 students who took Premiere’s course. Their study concluded that students who took Goldstar’s course scored significantly higher than students who took Premiere’s course. Is Goldstar justified in its claim that its SAT preparation course is superior to Premiere’s course?

x. The evidence strongly supports this claim.
y. The evidence contradicts this claim.
z. The evidence is not sufficient to support or contradict this claim.

Please think about the quantitative and scientific reasoning test that you are about to complete. Mark the answer that best represents how you feel about each of the statements below.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

1. Doing well on this test is important to me. (Importance)
2. I am not curious about how I will do on this test relative to others. (Importance*)
3. I am not concerned about the score I will receive on this test. (Importance*)
4. This is an important test to me. (Importance)
5. I would like to know how well I do on this test. (Importance)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>Disagree</td>
<td>Somewhat Disagree</td>
<td>Neutral</td>
<td>Somewhat Agree</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

6. I think I am up to the difficulty of this test. (Self-Efficacy)
7. I probably won’t manage to complete this test. (Self-Efficacy*)
8. I think everyone could do well on this test. (Self-Efficacy)
Appendix B

Students saw the following test instructions after completing the first 22 items (subtest1) on NW9. The instructions for items administered after the second 22 items (subtest2) and the third 22 items (subtest3) differed changing the wording to “… completing the 1st third of the test” to “…completing the 2nd third of the test” and “…completing the last third of the test”. Bolded words at the end of each item were added in this appendix to indicate what construct each item represents. * indicates reverse coded items. Students did not see this.

Tests can induce different feelings. The following items pertain to feelings you may experience DURING a test.

These items ask about how you are feeling right now, after completing the 1st third of the test.

Please indicate how you feel AT THIS MOMENT.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly Agree</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

31. Doing well on this test is important to me. (Importance)
32. I am engaging in good effort throughout this test. (Effort)
33. I am not curious about how I did on this test relative to others. (Importance*)
34. I am not concerned about the score I receive on this test. (Importance*)
35. This is an important test to me. (Importance)
36. I am giving my best effort on this test. (Effort)
37. While taking this test, I could work harder on it. (Effort*)
38. I would like to know how well I did on this test. (Importance)
39. I am not giving this test my full attention while completing it. (Effort*)
40. While taking this test, I am able to persist to completion of the task. (Effort)
41. While taking this test, I am thinking about how poorly I am doing. (Worry)
42. I am anxious. (Worry)
43. While taking this test, I am thinking about items I could not answer. (Worry)
44. I am proud of how well I am doing on this test. (Pride)
45. I am angry. \textbf{(Anger)}
46. I am enjoying this test. \textbf{(Joy)}
47. Because I am bored, my mind wanders. \textbf{(Bored)}
48. I am very pleased with my performance on this test. \textbf{(Pride)}
49. I am fairly annoyed. \textbf{(Anger)}
50. While completing this test, I am happy. \textbf{(Joy)}
51. I am bored. \textbf{(Bored)}
52. I am proud of myself. \textbf{(Pride)}
53. Thinking about this useless test makes me irritated. \textbf{(Anger)}
54. For me, the test is enjoyable. \textbf{(Joy)}
55. I am so bored, I have trouble staying alert. \textbf{(Bored)}
References


*Anxiety, Stress and Coping, 17*, 287-316.


