Summer 2013

# Measuring motivation for coursework across the academic career: A longitudinal invariance study

Makayla Grays
*James Madison University*

Measuring Motivation for Coursework across the Academic Career:

A Longitudinal Invariance Study

Makayla Grays

A dissertation submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

in

partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Department of Graduate Psychology

August 2013

# Acknowledgments

I extend my sincere gratitude to Robin, Chris, Dena, Christine, Teresa, John, Kenn, Bo, Jerusha, Jeff, Dan, Anna, and Ashley for all of the ways they supported me in the completion of this dissertation.

Table of Contents

List of Tables

List of Figures

## Abstract

Students must be sufficiently motivated in order to achieve the intended learning outcomes of their college courses. Research in education and psychology has found motivation to be context-dependent. Therefore, students' motivation is likely to differ from one semester to the next according to which courses students are taking. However, there are also instances in which motivation levels may not change over time. In order to determine whether motivation for coursework changes across the academic career (and, if so, what variables may be related to that change), it is imperative to use a measure of motivation that is theoretically and psychometrically sound. In addition, the measure should function consistently over time—that is, the motivation measure must demonstrate longitudinal invariance. The purpose of this research was to investigate the factor structure and longitudinal invariance of a measure of motivation for coursework—the Expectancy, Value, and Cost Scale (EVaCS)—for incoming and mid-career college students. Study 1 examined the factor structure of the EVaCS and found support for a correlated three-factor model. The longitudinal invariance of this model was examined in Study 2, and results established the EVaCS to be an invariant measure of motivation for coursework across the two time points. An analysis of latent mean differences showed no significant overall mean changes in Expectancy and Value over time, but a statistically and practically significant increase was found for Cost ($p < .05$, $d = 0.46$). In addition to establishing the EVaCS as a structurally sound instrument, this research has implications for the measurement of motivation for coursework and the theoretical conceptualization of motivation.

# CHAPTER 1.

# INTRODUCTION

Students' motivation is widely acknowledged to be an important element of postsecondary success. The research literature contains numerous examples of the positive impact that student motivation has on a variety of outcomes, including learning and performance in college courses (for reviews see Eccles et al., 1983; Wigfield, 1994; Wigfield & Cambria, 2010; Wigfield & Eccles, 1992). In order for students to successfully achieve the intended learning outcomes of their courses, it is crucial that they be sufficiently motivated, which relies in large part on their perceiving value (e.g., importance, purpose, relevance) in their coursework (Eccles et al., 1983). To this end, research on students' motivation is crucial. The quality of such research will necessarily rely, in part, on the quality of the instrument used to measure motivation. The purpose of the present research is to examine the psychometric structure and longitudinal invariance of a scale used to measure students' motivation for coursework. Results from this research will inform whether the scale can be used to measure motivation at different time points throughout the academic career. If so, the scale may then be used to better understand whether, how, and why motivation changes as students proceed through college.

## Motivation for Coursework across the Academic Career

Ideally, students would be highly motivated in all of their coursework throughout college. However, academic motivation is context-dependent in that students have different levels of motivation for different kinds of coursework (Bong, 2001; 2004). Therefore, it is quite unlikely that students will be equally motivated across all of their courses. Rather, motivation will differ according to the set of courses that students are taking at a particular point in time. In addition, because students enroll in a different set of courses each semester, their academic motivation is expected to vary from one semester to the next depending on which courses they are taking.

There are several ways in which coursework differs across the academic career. For instance, it typically becomes more rigorous as students move from introductory- to upper-level courses. Another key difference is that the focus of course enrollment shifts from general education to major coursework. Students traditionally complete the majority of any general education requirements within the first few semesters of college. Over time, as general education requirements are fulfilled, students begin taking more coursework in their academic majors. This shift in the composition of the curriculum across semesters has implications for students' motivation given that students often hold different attitudes about these two types of coursework. From the students' perspective, whereas major coursework is generally perceived as a necessary and important component of the undergraduate curriculum, the same is not always true of general education (Boyer, 1987). A common criticism argued by students is that general education is a waste of time and money that prevents them from taking additional coursework in their majors (Perk, 2005; Pracz, 2011; Wade, 2013; Zavislak, 2012).

**Motivation for General Education**

One might think that student motivation has been extensively researched within the context of general education. However, little work has been carried out in this area, and scholarly writings on student motivation for general education have seemed more anecdotal than empirical. The most comprehensive studies to date were conducted two to three decades ago (e.g., Boyer, 1987; Gaff & Davis, 1981) and provided a mixed picture of students' attitudes. Both the Boyer (1987) and Gaff and Davis (1981) studies noted that college students placed a strong emphasis on career preparation—an emphasis that many would claim has persisted, if not increased, among today's students.

**The impact of future careers.** Johnston et al. (1991) write that "the common wisdom is that, concerned as undergraduates are with career preparation and study in a major field, there is little 'demand' for [general education]" (p. 183). Indeed, there is an abundance of research indicating that the number one reason for attending college is related to career preparation. In theory, a high concern for careers does not necessarily jeopardize other areas of the curriculum; students can be simultaneously motivated for both major and non-major (general education) coursework. However, many would argue that a postsecondary vocational focus occurs at the expense of general education. For example,

> Most students these days are motivated primarily by a wish to prepare for a career… Most also assume that one's choice of major follows automatically from that of a career: journalism implies journalism; business implies business. Therein lies their academic plan—and in their eyes, in many cases, the fundamental "irrelevancy" of general education, however nice they think it would be to become broadly educated. (Johnston et al., 1991, p. 192)

**The main issue: Value.** The "fundamental irrelevancy" phrase used by Johnston et al. (1991) characterizes what seems to be the primary motivational challenge in general education: students' lack of value for the curriculum. Coursework that students perceive as irrelevant to career preparation—the reason most attend college in the first place—has little value, and students might therefore have lower motivation for this coursework. In particular, the perceived lack of relevance to a student's long-term career goals is an illustration of what is known in the motivation literature as *utility value*. Utility value refers to an activity's perceived usefulness for an individual's short- or long-term goals (Wigfield & Eccles, 2000), and has been shown to predict student outcomes such as interest, course-taking, and performance (e.g., Hulleman, Durik, Schweigert, & Harackiewicz, 2008). Unlike general education coursework, major coursework does not suffer from such a lack of value; indeed, for many students, it epitomizes how the curriculum contributes to their utility value for career goals.

**Why Value for Coursework May Change**

Early on in their academic careers students take primarily general education courses, which gradually become replaced by courses in their academic major (Koljatic & Kuh, 2001). Some research has shown that students tend to value coursework in their major more highly than general education (Grays, Hulleman, & Barron, 2012). Therefore, one might conclude that students' value for their coursework will be lower in semesters that mainly consist of general education courses (i.e., early), and it will be higher in semesters that consists of more major coursework (i.e., later). Comparing how value for coursework changes across semesters may reveal an increase in value over time, due mainly to the changing composition of coursework over time.

**Why Value for Coursework May Not Change**

Although students may report higher value for major coursework than general education, students can be highly motivated for *both* types of coursework. One indication that this may be the case comes from the national Freshman Survey, conducted annually through the Higher Education Research Institute at UCLA.  Students were asked on the survey whether the opportunity “to gain a general education and appreciation of ideas” was a “very important” reason for their decision to attend college.  The majority (73%) of freshmen entering college in 2012 indicated that this was a “very important” reason (Pryor et al., 2012).  Although it seems logical that students’ reasons for attending college would be related to their motivation for college coursework, in which case one might conclude that most students *are* motivated for general education, this link has not been established through research.  In terms of how this may impact value for coursework over time, if first-year students are in fact interested in acquiring a broad appreciation of ideas, which the general education curriculum affords, they may not exhibit an increase in value.  This is because students would value the early general education experiences *and* (presumably) major coursework later in their academic careers.  Thus, their value for coursework would remain more or less the same over time even as the composition of coursework changes.

Another segment of students who may not exhibit an increase in value for coursework over time is students who enter college without a decided major and therefore are less able to say whether their coursework is relevant to a major or future career.  In comparison, students who have decided on a major are in more of a position to judge whether or not coursework is relevant to that major.  For example, if a course is a requirement or elective within the major, it is likely relevant; otherwise it is not, or is somewhat less relevant.  Decided and undecided students have been found to differ somewhat in their reasons for attending college: Decided students are more likely to attend college for career preparation, whereas undecided students are more likely to emphasize intellectual development (Baird, 1967, in Gordon, 2007). The fact that these students' reasons for attending college differ suggests potentially different attitudes toward general education. To a decided student who is focused on career preparation, general education may seem like an unnecessary annoyance in his college experience. Conversely, to an undecided student who is focused on intellectual development, general education may present an interesting and useful opportunity to achieve her goal of intellectual development while at the same time helping her to select a major through exploration of the curriculum. In fact, one of the reasons students may be undecided is that they have many diverse interests (Cuseo, 2005). Therefore, if undecided students value their early general education experiences as well as later coursework in their major, they may display no difference in their level of value for coursework across semesters.

Clearly there are many possible reasons why motivation—particularly value—may or may not change as students progress through college. One reason is that the composition of coursework (major vs. general education) changes across semesters. Other reasons are related to students' attitudes toward general education coursework, which may also be influenced by their reasons for attending college (e.g., occupational emphasis vs. acquiring broad knowledge) and how decided they are about their future careers. To complicate the issue even further, many of these variables that may influence change in motivation can also change over time. For example, an incoming freshman who initially anticipated taking a broad, liberal arts-like approach to his postsecondary studies may find that his priorities begin to shift to a more occupational focus as he experiences competing demands for his time in college. In addition, students may experience more uncertainty about their future careers after being exposed to a wider range of options or to the realities of study in their intended fields (e.g., the demands of courses in pre-professional areas of study like medicine).

## Measuring Motivation for Coursework

In order to assess whether students' motivation for coursework changes over time, a suitable measurement instrument is needed. Several different scales, based on a variety of motivational theories, have been applied in higher education settings. The expectancy-value framework (Eccles et al., 1983) is especially appropriate for the present research for several reasons. This theory incorporates many key constructs from other theories—such as self-efficacy, achievement goal, and interest theories—into its two primary constructs, *expectancy* and *value* (Barron & Hulleman, 2006); these constructs are discussed in detail in Chapter 2. In particular, this framework includes utility value as one of several types of value, described above as being particularly pertinent to the issue of students' motivation for general education and major coursework. Because it highlights the value construct and also provides a broadly encompassing framework for understanding motivation, expectancy-value represents an appropriate framework for developing a measure of students' motivation for coursework.

### The Expectancy, Value, and Cost Scale (EVaCS)

The EVaCS (Barron & Hulleman, 2010) is a motivation instrument that was developed based on the expectancy-value theory of motivation. Items were written to measure one of three general constructs: *expectancy*, an individual's belief about how well he will do on an upcoming task; *value*, the reason(s) an individual engages in or attempts to succeed at an activity; and *cost*, the extent to which successfully engaging in an activity is constrained by other factors. Theoretically, the EVaCS appears to be a sound and promising tool for measuring motivation for college coursework. However, there has been limited psychometric work conducted on the EVaCS. For example, a version of this instrument was developed to measure expectancies, values, and costs within the context of a single course for high school (Getty, Hulleman, Barron, Stuhlsatz, & Marks, 2013) and college students (Flake, Barron, Hulleman, Lazowski, Grays, & Fessler, 2011; Kosovich, 2013). However, to date, no research has examined the EVaCS for college students across all their courses in a given semester, and more research on the scale is needed before it can be used to make inferences about students' motivation over time.

## Purpose of the Research

The purpose of this research was to evaluate the psychometric structure of a motivation scale and its use for making inferences about how students' motivation for coursework changes across the academic career. This was achieved through two studies. The first study focused on the structure of students' motivation for coursework as measured by the EVaCS. The second study focused on whether the EVaCS functioned similarly in incoming and mid-career student samples, and if so, whether there were mean-level differences in motivation for coursework over time.[1]

### Study 1: Confirmatory Factor Analysis

**Does the hypothesized three-factor model fit the EVaCS data better than the alternative models tested?** A series of theoretically-plausible models were tested using confirmatory factor analysis (CFA) to evaluate the model-data fit of the EVaCS. The

[1] The type of invariance under investigation here is *longitudinal* measurement invariance, meaning that the same sample provides data at different time points, as opposed to *multiple-group* measurement invariance, for which independent samples provide data. When references are made to "incoming and mid-career" students throughout the present research, this means the same group of students assessed twice—first as incoming students, and again as mid-career students—not two independent groups of students.

best-fitting model, as determined by various fit indices and model parsimony, is the one which best represents the relationship among any latent constructs believed to be driving students' responses to the EVaCS items. Four a priori models were examined: (a) a unidimensional model, (b) a correlated two-factor model, (c) a bifactor model, and (d) a correlated three-factor model. These models are shown in Figures 1-4 and are described in Chapter 3.

**Study 2: Measurement Invariance and Latent Mean Differences**

**Does the EVaCS exhibit longitudinal measurement invariance for incoming and mid-career students?** The best fitting model from Study 1 was used to test for longitudinal measurement invariance (configural, metric, and scalar) across incoming and mid-career students. Measurement invariance means that the model's parameters, if fixed from one time point to the next, produce adequate model-data fit when data from both time points are fit simultaneously. It is important to establish measurement invariance if any longitudinal comparisons (e.g., mean differences) are to be made accurately (Vandenberg & Lance, 2000). Without invariance, it would be unclear whether any differences (or lack of differences) over time are due to true differences (or lack thereof) in the latent constructs or rather due to the instrument functioning inconsistently across time points.

**Do the latent means of motivation for coursework differ for incoming and mid-career students?** If measurement invariance is established, it then becomes possible to interpret differences in incoming and mid-career students' motivation as measured by the EVaCS. The motivational construct of primary interest in this study is value, but given the many different variables (not examined in this research) that may influence motivation for coursework at either time point, no hypotheses were made about how value might change.  In addition, no hypotheses were proposed about differences across time for the expectancy and cost constructs.

# CHAPTER 2.

# REVIEW OF THE LITERATURE

This chapter provides a background for understanding how and why students' motivation for coursework may change across the academic career. It begins with a discussion of expectancy-value theory, then proceeds into how this theory applies in the contexts of major and general education coursework. The value construct, particularly the *utility* subtype, is frequently emphasized in light of the fact that student motivation for college is largely focused on future careers. Consideration is given to how motivation for coursework may be different for undecided and decided students, which leads into a discussion of why differences in motivation for coursework may or may not be observed for incoming and mid-career students. Finally, the instrument at the center of the present research is discussed.

## Expectancy-Value Theory of Motivation

Among the many different theories of motivation, one of the most widely researched and well-known is expectancy-value theory. According to this theory, an individual's belief about how well she will perform an activity and the extent to which she values the activity influence her choice of, persistence at, and performance on the activity (Wigfield & Eccles, 2000). As applied to an academic context, students will be most motivated for an educational activity (e.g., a course or specific course unit) when their expectations for success and value for the activity are high. The expectancy-value model developed by Eccles and colleagues (Eccles et al., 1983; Wigfield & Eccles, 2000) is the most researched model representing this theory with regard to academic achievement (Conley, 2012).

**Expectancy**

Expectancy is an individual's subjective appraisal of potential success at an activity; it is represented in the self-posed question, Can I do this activity? Eccles and Wigfield (1995, p. 215) note that "expectations for success…have been assigned a central role in almost all cognitive theories of motivation." In his theory of achievement motivation, Atkinson (1957) defined expectancy as the proportion of individuals who are successful at a particular activity—i.e., an objective mathematical probability of success. This definition of expectancy neglected the subjective aspect of individuals' success appraisals which is likely to influence their motivation as well, particularly because in many situations the mathematical probability of success at a given activity is unknown. For instance, an academic activity may have an objective success probability of .50 (a student is as likely to succeed at the activity as he is to fail), but a student who perceives himself as highly-able may judge his probability of success to be greater than .50. Conversely, a student who regards himself as less able may assume a lower probability of success for the same activity.

Eccles et al. (1983) incorporated the subjective element into their definition of expectancy, which is an individual's belief about how well he or she will do on an upcoming task. There are two main components underlying the concept of expectancy. The first is an individual's *ability beliefs,* or perception of his current competence in a particular domain. The second component is *expectancy*, or an individual's expectation of success on a specific upcoming activity. The conceptual distinction between these two components is that ability beliefs are focused on present ability whereas expectancies are future-oriented (Wigfield & Eccles, 2000). However, ability beliefs and expectancy have a high empirical relationship (Eccles & Wigfield, 1995), and items representing each are often used to represent a general expectancy construct. Eccles and Wigfield (2002) note that these two components are likely to be indistinguishable in actual achievement contexts. Example expectancy items are shown in Table 1.

Expectancy is related to task difficulty; however, task difficulty has been found to be distinguishable from ability beliefs (Eccles et al., 1983; Eccles & Wigfield, 1995). Eccles and Wigfield (1995) measured students' perceived task difficulty using items that pertained to task difficulty and required effort. They reported a negative relationship between task difficulty and ability beliefs, meaning that students who considered themselves competent in performing an activity also viewed the activity as easier and requiring less effort. Because it is theoretically distinct from expectancy, perceived task difficulty is not always included on expectancy-value measures of motivation.

**Value**

Whereas expectancy focuses on the question, *Can* I do this activity?, value focuses on the question, Do I *want* to do this activity? Task value (often shortened to just *value*) can be thought of as the reason(s) an individual engages in or attempts to succeed at an activity. The modern conception of value bears little resemblance to Atkinson's (1957) concept of incentive value, which is the pride an individual feels after accomplishing a task. Atkinson defined incentive value in terms of expectancy (the probability of success), mathematically expressed as

$$\text{incentive value} = 1 - P_{\text{success}}$$

In this formulation, the incentive value of a task is completely defined by one's probability of being successful at the task. Therefore, an inverse relationship exists between expectancy and value: Tasks with a lower probability of success are more valued—that is, they lead to a greater sense of pride and feelings of accomplishment. Although Atkinson's formula may be somewhat useful in representing why people pursue or persist at the most difficult tasks (because they should result in immense pride), it does not account for reasons other than pride which can also motivate people to achieve tasks, regardless of a specific task's probability of success.

An activity may be considered valuable for a number of reasons. Eccles and colleagues' conception of task value elaborates on earlier expectancy-value models (Atkinson, 1957) by more fully explicating different types of value. Specifically, four types of task value are posited: importance, or attainment value; interest, or intrinsic value; usefulness, or utility value; and cost (Eccles et al., 1983; Wigfield & Eccles, 1992; Wigfield & Eccles, 2000). *Attainment value* is the importance of doing well at a particular activity. For example, a student might value succeeding in his college courses in order to preserve his identity as a capable individual. *Intrinsic value* is the inherent enjoyment one gains from doing an activity. A student who enrolls in a religion course due to a personal interest in the topic is exhibiting intrinsic value. *Utility value* is how useful an activity will be for an individual's short- or long-term goals. The perceived usefulness of a particular course to a student's future occupation, for instance, will influence her utility value for that course. *Cost* is the extent to which successfully engaging in an activity is constrained by other factors. Examples of cost include amount of effort required to be successful, psychological factors (e.g., anxiety, stress, fear of failure), and the inability to engage in other valued activities (loss of valued alternatives). Unlike attainment, intrinsic, and utility value which increase task value, cost is thought to decrease task value (Eccles & Wigfield, 1995). Example value items are shown in Table 1.

Although attainment, intrinsic, and utility value are considered theoretically distinct, they tend to have moderate to high positive correlations (e.g., Conley, 2012; Eccles & Wigfield, 1995; Luttrell et al., 2010; Trautwein et al., 2012). As a result, items representing these three value subtypes are sometimes combined to form a 'general' value scale (e.g., Flake et al., 2011). Correlations between cost and the other three value subtypes tend to be smaller, particularly for cost and utility value (Trautwein et al., 2012), although the pattern of correlations varies considerably across studies.

**Cost**

Of the four value subtypes proposed by Eccles and colleagues, cost has been the least studied (Wigfield & Cambria, 2010), although it has recently drawn greater attention from motivation researchers (Flake, 2012). It is worth noting that cost was initially introduced by Eccles et al. (1983, pp. 93-95) as something that influences value rather than a specific type of value. Since then, value investigations have focused more on attainment, intrinsic, and utility value, and their relationships with cost has been largely neglected. The varying degrees of correlation that have been observed between the three primary value subtypes and cost—moderate negative (Flake et al., 2011), small positive (Conley, 2012), and small-to-moderate positive (Luttrell et al., 2010; Trautwein et al., 2012)—is likely related to the items' differential emphasis of cost content and how the items are framed or analyzed. For example, in terms of content, Trautwein et al.'s (2012) cost scale focused exclusively on loss of time, while Luttrell et al.'s (2010) scale emphasized psychological contributors to cost such as anxiety, fear, and worry. Cost items tend to be framed such that selecting a high response option indicates high cost—e.g., selecting ‘5’ on a 1 (*strongly disagree)* to 5 (*strongly agree*) scale for “I have to give up a lot to do well in math.” However, in their analyses, Luttrell et al. (2010) reverse-scored the cost items, and Trautwein et al. (2012) reported results for low cost. Such issues contribute to confusion about cost’s appropriate place within the expectancy-value framework.

Some have argued that cost is most appropriately treated as distinct from value rather than as a specific subtype (Barron & Hulleman, 2010), and studies have demonstrated how cost can be differentiated from general and specific types of value (Flake et al., 2011; Getty et al., 2013; Kosovich, 2013). The theoretical implication for modeling cost separately from value is that high cost may not have a direct negative effect on value alone but rather on motivation more generally (i.e., on value *and* expectancy).

**Relationships between Expectancy, Value, and Cost**

Studies have shown that expectancy and attainment, intrinsic, and utility value generally have moderate to high positive correlations. Eccles and Wigfield (1995) offered suggestions as to why these relationships exist. Regarding attainment value,

> A…positive association between [expectancy] and attainment value (perceived importance) seems likely to the extent that individuals are interested in maintaining a positive self-image. One effective way to maintain one’s self-esteem is to rate as very important those activities that one is most confident about succeeding at and to rate relatively less important those activities one is least confident about succeeding at. (p. 217)

Regarding intrinsic value,

> Individuals should come to like or enjoy (intrinsically value) those activities at which they have done well at in the past and are reasonably confident of being able to succeed. Conversely, individuals should come to dislike those tasks that they have done poorly at in the past. (p. 217)

Their hypothesis about the relationship between utility value and expectancy was somewhat tentative, though still supportive of a positive correlation.

> Because the perceived utility of any particular task is determined by its links to goals and activities that are extrinsic to the task, utility can be influenced by a wide range of things… Given these other influences on utility value, we predicted that the positive links between [expectancy] and [utility value] will be weaker than the links between [expectancy] and attainment value and interest. (p. 217)

These explanations are supported by many examples of positive correlations, but the theoretical and empirical distinctiveness of expectancy and value and the absence of extremely high correlations is also important to recognize. Although high expectancy is typically associated with high value, it is also possible for students to have high expectancy for activities which they do not highly value. Likewise, they may have high value for activities at which they do not expect to do well.

The relationship between expectancy and cost remains unclear, perhaps due to fewer studies examining cost, or to inconsistencies in how cost is measured. Conley (2012) reported virtually no relationship between competence beliefs and cost, $r = .02$. Trautwein et al. (2012) found a strong positive association, $r = .75$, between expectancy and low cost. A possible explanation for the difference in results is that Conley's cost items utilized general cost terminology (e.g., "I have to give up a lot to do well in math") while the items used by Trautwein et al. were focused specifically on the loss of time (e.g., "I'd have to invest a lot of time to get good grades in mathematics"). It makes sense that cost would be negatively correlated with expectancy, particularly when cost items' content is effort-related. Easy tasks require less effort to accomplish; thus, for an activity that is perceived as difficult relative to one's capabilities (low expectancy), more effort may be necessary to be successful (high cost). Using a general cost scale with some effort-related items, Flake et al. (2011) observed a moderate negative correlation with expectancy, $r = -.30$.

**Antecedents and Consequences of Expectancy and Value**

The positive relationship between expectancy and value may be attributed to the fact that they share many of the same antecedents. In their full model of achievement motivation, Wigfield and Eccles (2000) illustrate how expectations of success and task value are both directly influenced by self-schemata, short- and long-term goals, self-concept of one's abilities, and perceptions of task-demands. For instance, a college student's future career goals are believed to impact her expectancy and value for college coursework. Expectancy and value are also impacted through indirect and interactive effects from such things as actual and perceived stereotypes, aptitude, previous achievement-related experiences, and interpretation of experiences (causal attributions, locus of control). Therefore, a student's expectancy and value for her college coursework might also be related to things like standardized test scores and academic performance in high school.

What makes the study of students' motivation and its antecedents so important is the fact that expectancy and value are both related to achievement outcomes. Researchers have typically found expectancy to be the stronger predictor of performance, whereas value is more strongly related to task choice, effort, and persistence (Eccles, 1983; Wigfield, 1994; Wigfield & Eccles, 1992). Although much of the research supporting the expectancy-value model's theorized relationships has studied children and adolescents (Wigfield, 1994), applications have more recently been made in higher education settings (e.g., Flake et al., 2011; Hulleman et al., 2008; Luttrell et al., 2010). For example, Flake et al. (2011) found value to be a significant predictor of interest in an introductory psychology course, and expectancy was the primary predictor of course grade.

### A Major Source of Undergraduates' Motivation: Future Career Goals

Much research conducted with undergraduates within the last decade has reported that students' reasons for attending college are now largely career-focused—e.g., to get a job, to prepare for a career (Bui, 2002; Gordon & Steele, 2003; Kennett, Reed, & Lam, 2011; Phinney, Dennis, & Osorio, 2006). Results from the most recent (2012) Freshman Survey, administered annually by the Higher Education Research Institute at UCLA, showed that "the ability to get a better job" was the top reason for attending college cited by entering freshmen: 88 percent of students—an all-time high—said this was a "very important" reason for attending college. It has been argued that the rising costs of college and the likelihood of loan debt may necessitate students' pragmatic, occupational approach to their studies (Sander, 2013). As Johnston et al. (1991) note, "the unprecedented expense of [a college] education also gives [students] a powerful incentive to fix their primary attention on rewarding careers" (p. 190).

**Motivation for Career-Related Coursework**

With career preparation topping the list of reasons why students choose to attend college, it follows that an occupational emphasis may influence their achievement motivation while in college. Johnston et al. (1991) note that "most students these days are motivated primarily by a wish to prepare for a career" (p. 192). Attending college with the goal of preparing for a future career is a clear illustration of utility value: Students who have a career goal in mind should value the coursework and activities which they perceive as relevant to their future careers. Students who have decided upon a career may have done so in part due to their intrinsic interest for a specific subject or area—e.g., a student with an intrinsic interest in music aspires to become a professional musician and chooses to major in music performance while in college. In this case, the student should have high intrinsic and utility value for his music courses. Students who see the importance of being successful in their career-related coursework should also have high attainment value. For instance, a student who views being an effective accountant as part of his professional identity will probably place great importance on succeeding in the courses he takes as part of the accounting major. Grays et al. (2012) surveyed incoming college freshmen about their general value for coursework in their major and found that, on average, students rated their value at a mean of 5.71 (*SD* = .58) on a 6-point scale.[2] Students rated their intrinsic value for this coursework at a mean of 5.61 (*SD* = .68).[3]

Future career goals should play a substantial role in influencing students' value for career-related college coursework. Additionally, to the extent that students pursue careers in fields in which they have high ability beliefs and expectations for success, their expectancy should be high as well. Grays et al. (2012) found that incoming college freshmen reported fairly high expectancy for coursework in their major. On average, students rated their expectancy at a mean of 5.39 (*SD* = .70) on a 6-point scale.[4] Although students' high expectancy for coursework in the major is anticipated, it may be tempered by a more persistent belief in the difficulty of courses within their major, or the difficulty of college coursework in general. However, a recent study found that to high school students, college is generally not viewed as a challenging experience: Only 11 percent of students said that they expected college to be difficult (Adams, 2012).

## General Education

Students with a particular career in mind may have high expectancy and value for coursework in their major, but often students do not begin taking courses in their major until the second or third year of college. Instead, at most higher education institutions, the first two years of undergraduate study are largely devoted to the general education curriculum (Koljatic & Kuh, 2001). General education plays an important role in

[2] General value for the major was measured by the item, "The topics and skills taught in my major courses are important to me" (1 – *strongly disagree,* 6 – *strongly agree*).
[3] Intrinsic value for the major was measured by the item, "My major courses interest me" (1 – *strongly disagree*, 6 – *strongly agree*).
[4] Expectancy for the major was measured by the item, "I can do well in my major courses" (1 – *strongly disagree*, 6 – *strongly agree*).

providing students a liberal education, which the Association of American Colleges and Universities (AAC&U, n.d.) describes as:

> …an education that exposes students to a wide breadth of courses, perspectives, and educational experiences designed to equip them with the essential skills and learning necessary to thrive and succeed throughout their lives. A liberal education prepares students to deal with complexity, diversity, and change, and entails study across many fields, as well as in-depth study in a specific area of interest. A liberal education helps students develop a strong sense of personal and social responsibility—important in all spheres of life.

The terms *general education* and *liberal education* have often been used interchangeably (Brint et al., 2009; Mulcahy, 2009); however, a distinction is necessary because general education is only a component of liberal education. Liberal education as defined by the AAC&U is comprised of general education coursework ("study across many fields"—i.e., breadth) *and* major-specific coursework ("in-depth study in a specific area of interest"—i.e., depth). Another way general and liberal education can be distinguished is that whereas the goals of liberal education are fairly consistent across institutions, the particular manner in which they are achieved—e.g., through the general education curriculum—often varies from one institution to the next (e.g., Brint et al., 2009). The institution-specific aspect of general education is noted in the following definition from Jones, Hoffman, Ratcliff, Tibbets, & Click (1994, in Bourke, Bray, & Horton, 2009, p. 219):

> General education is frequently taken to mean the collection of experiences crafted by the institution to provide students with a breadth of learning

> experiences and a broad knowledge base that sharpen students' problem-solving, interpersonal, and oral and written communication skills, as well as their cultural and linguistic literacy.

General education can therefore be understood as an institution's curricular mechanism for achieving the breadth-related goals of liberal education. Moreover, it represents something of a shared academic experience—"that portion of the curriculum studied by all students, regardless of their academic major or intended career" (Gaff, 1989, in Reardon, Lenz, Sampson, Johnston, & Kramer, 1990, p. 2).

General education coursework constitutes roughly one-third of the credits students will accumulate in college (Brint et al., 2009; White & Cohen, 2004). General education is most commonly structured as a distribution model, which emerged in the early to mid-1900s as a way to achieve greater coherence in the college curriculum (Bourke et al., 2009; Brint et al., 2009). Under the distribution model, students choose which courses to take within each general education area (e.g., Arts and Humanities), with the number of courses and any additional restrictions prescribed by the institution. Gaff and Wisescha (1991) reported the typical general education distribution is comprised of the following: four humanities courses; three social sciences courses; two natural sciences courses; two writing courses; one mathematics course; one fine arts course; and additional coursework such as foreign language, physical education, speech, computer literacy, or quantitative reasoning. Modern general education curricula draw heavily from traditional liberal arts disciplines (e.g., history, languages, literature, and philosophy) which can be contrasted with more vocational-oriented fields (e.g., business, education, engineering, nursing and other health professions).

**Liberal Arts vs. Vocationalism (or *Education* vs. *Training*)**

American higher education is, in a sense, rooted in vocationalism. Many of the earliest institutions were founded to prepare young men for leadership roles in the church, medicine, or law through classes in Greek, Latin, mathematics, and moral truths (Fuhrmann, 1997, in Bourke et al., 2009). Thus, there was no division between general and specialized (vocational) education (Rudolph, 1977, in Boning, 2007). The distinction is now apparent as college catalogs and websites clearly delineate the number of credits students need to fulfill general education and major requirements. Most scholars assert that colleges and universities have the responsibility to provide both a general education and vocational specialization, often characterized as *breadth* and *depth*, respectively. However, there is disagreement regarding the relative emphasis that each should receive. A study conducted by the Pew Research Center (2011) found that 47 percent of the general public say the main purpose of college is to teach work-related skills, 39 percent say the main purpose is to help students grow personally and intellectually, and 14 percent placed equal importance on both purposes. College graduates were more likely to emphasize intellectual growth, while those without a college degree were more likely to emphasize career preparation (Pew Research Center, 2011).

The proper focus on vocational preparation is debated within institutions too, as many campuses are "torn between careerism and the goals of liberal learning" (Boyer, 1987, p. 105). The offering of vocational majors (e.g., business administration, nursing, teacher education) attracts many prospective students, which is ultimately beneficial for institutional enrollments. Greater access to and demand for postsecondary education has prompted the creation of more of these majors, perhaps most notably the proliferation of business-related majors (Boyer, 1987). According to the National Center for Education Statistics, of the over 1.6 million bachelor's degrees awarded in 2009-10, most were in a vocational major, the largest share being in business (21.7 %). Just 2.8 percent were in the liberal arts and sciences, general studies, and humanities (Aud et al., 2012).[5] Still, some in higher education have expressed mild to severe opposition toward vocational offerings that, if narrowly focused or improperly emphasized, can interfere with a well-rounded liberal education. Boyer (1987) described how "many faculty members, especially those at liberal arts colleges, voiced the opinion that it is inappropriate for colleges to offer majors that are primarily 'vocational'" (p. 108). Some have questioned whether vocational training belongs in four-year institutions at all (Selingo, 2013) while others refuse to even acknowledge it as a form of education. For instance, Côté and Allahar (2011) argue that "while one may be *trained* in engineering, one can only be *educated* in the liberal arts and sciences" (p. 15, emphasis in original).

Despite ongoing dispute over appropriate emphases in the undergraduate curriculum, employers seem far from discounting general education in favor of vocational training. A national survey of employers found that a majority believe colleges should place even more emphasis on general education outcomes, such as effective oral and written communication, critical thinking and analytical reasoning, than they currently do as these are deemed vital for workplace success (Hart Research Associates, 2010). Thus, employers seek workers who possess not only discipline-specific knowledge (as is gained through specialized study in a major) but also the type

[5] The National Center for Education Statistics' "liberal arts and sciences, general studies, and humanities" classification does not include some fields often considered liberal arts disciplines, such as social sciences and history (10.5% of degree recipients in 2009-10), psychology (6.1%), and English language and literature/letters (3.2%) (Aud et al., 2012).

of skills that the general education curriculum is intended to help students develop. Unfortunately, there are indications of employer dissatisfaction with recent graduates' communication, problem solving, and decision-making skills (Fischer, 2013). What is ultimately important is that students recognize the value of their general education coursework, if not for its own sake in rounding out their academic experience, then for how it can support their career ambitions.

**Student Attitudes**

Student attitudes toward general education can offer important insight into development of the skills sought-after by employers. However, students are "the most neglected audience among the various participants in general education" (Reardon et al., 1990, p. 5). Few examples of research on student attitudes toward general education exist. The most recent national Freshman Survey (Pryor et al., 2012) noted that 73 percent of freshmen said that gaining a general education and appreciation of ideas was a "very important" reason for attending college, up from 66 percent a decade earlier. Boyer (1987) noted that undergraduates reported higher enthusiasm for general education courses in which connections are made to contemporary issues, and Gaff and Davis (1981) reported that "students are more likely to support general education if it is formulated to include 'personal and interpersonal' dimensions" (p. 188).

Despite these findings, it is not uncommon for undergraduates to view the general education curriculum as contrary to what many consider to be their primary reason for attending college—career preparation. Students often view general education requirements as "something to 'get out of the way' [rather than] an opportunity to gain perspective" (Boyer, 1987, p. 84) before they are able to focus on studies within their major (White & Cohen, 2004).

> Students overwhelmingly have come to view general education as an irritating interruption—an annoying detour on their way to their degree. They all too often do not see how such requirements will help them to get a job or live a life. (Boyer, 1987, p. 102)

Johnston et al. (1991) also discuss how students often do not perceive value in their general education coursework.

> General education seems for many undergraduates an imposition rather than a welcome opportunity for intellectual challenge and growth… Too few students seem to understand its purposes and importance, and too few recognize the possibility of its being…relevant to their interests and aspirations. (p. 182)

Although formal research may be lacking, many students have made their attitudes toward general education evident through other outlets. The following quotes are from recent articles in college student newspapers on the topic of general education coursework. In each instance, the student comments on the perceived irrelevance of general education courses to his or her future career.

> The average history major won't be so moved by an introductory course in psychology to change their major, nor will it be particular relevant to their future career aspirations or major coursework… For the most part, general electives are a waste of time, and consequently, money. Universities would be better off…if they abolished [general education] requirements for students and allowed us to

> instead focus on what we came to school to study. (Wade, 2013, in *The Commonwealth Times*, Virginia Commonwealth University)

Several of Wade's (2013) remarks regarding a stronger emphasis on major coursework are also made by Pracz (2011), who notes how students' approaches to major and general education coursework differ.

> General education courses should not be required for college students… Many of these classes are rather useless since they probably have absolutely nothing to do with one's major… For the most part, students should be able to focus solely on their major during their time at college. The fact is that most students don't approach [general education] courses with the same sort of devotion as they do with classes they see as being useful… College is so expensive that it is hard to justify taking classes that ultimately do not help you in your career. (Pracz, 2011, in *Northern Star*, Northern Illinois University)

Like Wade (2013) and Pracz (2011), Zavislak (2012) comments on how general education requirements take time away from coursework in the major.

> General education requirements take away from students who wish to devote more time to and take more classes in their respective majors… Instead of requiring students to devote a certain number of blocks to subject areas that they do not like or areas that are not remotely tangent to their future careers, colleges should allow students to pursue whatever courses they wish. (Zavislak, 2012, in *The Cornellian*, Cornell College)

In each article, these students argue that students should not be required to "waste" their time or money completing coursework in subjects that they find neither interesting nor relevant to their majors or future careers. Apart from concluding that these students represent a particular minority of today's undergraduates, it is difficult to reconcile these students' perspectives with the Higher Education Research Institute's Freshman Survey finding that a majority of students seem eager to gain a general education upon entering college. In summarizing the literature on student views of higher education and general education specifically, Reardon et al. (1990) point out that while students seem to have goals related to specialized and general studies, "they do not have a very clear strategy for how to integrate these two goals, which are sometimes viewed as antithetical within higher education" (p. 13). Therefore, although students may indeed report interest in both career preparation and gaining a general education (as the latest Freshman Survey shows), when faced with the choice of where to direct their attention and effort, they may select those activities and experiences which have more direct vocational relevance, potentially at the expense of general education. If students are to acquire the knowledge and skills intended through the general education curriculum, they must see the value in this coursework and not view it as irrelevant to—or worse, interfering with—their career preparation.

**Research on Motivation for General Education**

To date, students' motivation for the general education curriculum has received minimal attention as an area of study. Johnston et al. (1991) note that "little has been done to identify and understand student perceptions regarding general education, much less reshape and harness them on its behalf" (p. 182). Despite a lack of extensive empirical research in this area, concern over students' lessened motivation for the general education curriculum does not appear to be unfounded. Writing specifically about problems with the distribution model of general education, the AAC&U (1994) reported "students generally did not see the utility of studying general education materials and thus lacked motivation or interest in mastering the traditional liberal arts subject matter" (in Warner & Koeppel, 2009, p. 243).

As one of few examples of research in this area, Miller and Sundre (2008) compared first-year students' and sophomores' motivation for general education coursework to their motivation for overall coursework in a semester.[6] It is important to note that general education and overall coursework were not mutually exclusive in their study. That is, students' overall coursework was likely to include general education courses. The authors found that first-year students and sophomores were less motivated to learn the material in their general education courses than in their overall coursework, and the disparity was larger for sophomores. They also found that first-year students and sophomores reported higher work-avoidance (exerting minimal effort) for general education than overall coursework, and again, the disparity was larger for sophomores. Miller and Sundre's findings suggest that students enter college with lower motivation for general education relative to other types of coursework, and that their motivation to learn in general education courses may decline as students proceed through the general education curriculum.

The measure used by Miller and Sundre (2008) asked students to report on their motivation for general education coursework as a whole, but in fact students are likely to hold different attitudes toward different components of the general education curriculum (Petrosko, 1992). For example, a student may be highly motivated for general education

[6] Miller and Sundre's (2008) questionnaire was based on achievement goal theory (Finney, Pieper, & Barron, 2004; Pieper, 2003), and their results emphasized mastery-approach and work-avoidance goals.

coursework in the natural sciences but less motivated for social science coursework. Grays et al. (2012) hypothesized that students' expectancy and value would vary across the different components of the general education curriculum and would be highest for coursework that was most closely related to their declared major. Results from their survey of incoming freshmen supported both of these hypotheses. Although the Grays et al. study illustrates how students' motivation differs depending on the disciplinary focus of general education courses, their measure was basic: A single item was used to assess expectancy and value within each component of the curriculum.

**Comparing Motivation for General Education and Major Coursework**

The issue of students' motivation for general education relative to their motivation for coursework in the major is especially intriguing given the strong occupational focus reported by today's students. The tendency might be to assume that the higher a student's motivation for her major is, the lower her motivation will be for general education (i.e., there is an inverse relationship). However, this has not been demonstrated empirically and might not be the case. It is possible for students to have high motivation for *both* the coursework in their majors and general education coursework; indeed, this is a most desirable motivational situation. Furthermore, even if a motivational discrepancy exists between general education and major coursework, students' motivation for general education may still be considered sufficiently high in an absolute sense, as Johnston et al. (1991) acknowledge:

> In study after study[,] the collective level of support for general education is only moderately below that for goals relating to career preparation, and high enough to

> indicate that many, and probably most, students are acknowledging major goals in both categories. (pp. 185-186)

Grays et al. (2012) sought to explicitly examine and compare students' motivation for general education curricular areas and coursework in their major. Incoming freshmen completed a brief survey of their motivation for each of the university's five general education areas (e.g., Arts and Humanities) and major, if they had officially declared one. The survey items were based on expectancy-value theory. Students responded to the items "I can do well in my [area # or major] courses" (expectancy) and "The topics and skills taught in my [area # or major] courses are important to me" (value) on scale of 1 (*strongly disagree*) to 6 (*strongly agree*). Across all entering freshmen with declared majors, the average motivation reported for coursework in their major was fairly high: mean expectancy = 5.39, mean value = 5.71. Motivation for the five general education areas was lower by comparison: mean expectancy = 4.59 to 5.12 ($d$ = -0.81 to -1.11), mean value = 4.41 to 4.93 ($d$ = -0.99 to -1.27). Grays et al. had two interesting and relevant findings. First, students reported reasonably high expectancy and value for *both* general education and major coursework, albeit lower for general education. Thus, in an absolute sense, students' responses were not altogether alarming as incoming freshmen seemed to have sufficient motivation for general education coursework. Second, the discrepancy between motivation for general education and motivation for the major was larger for value than expectancy. In other words, students' expectancies for general education and their major were more similar than their value for general education and their major, suggesting that any motivational challenges in general education are likely to be due to value more so than expectancy. This finding, as well as previously presented depictions of students' attitudes toward general education, point squarely to value as the most important construct in describing student's motivational deficits in general education. Low motivation for general education seems more likely to occur when students do not perceive the material as relevant to their future careers.

## Undecided Students

If undergraduates' motivation for general education is largely a function of value—particularly utility value—then what can be said about students who have not yet declared a major or decided on a future career? So far, the discussion of utility value has emphasized general education and major coursework's relevance for students' future careers (a long-term goal). However, utility value can also be understood as usefulness for accomplishing a short-term goal, like selecting a major. Thus, for students who have not yet decided on a major, general education may have high utility value because it presents an opportunity to explore potential interests in a variety of areas before committing to a field of study.

Despite the fact that upon entering college, only about 8 percent of freshmen report being undecided in terms of a major and 13 percent undecided in terms of a future occupation (Pryor et al., 2012), there is evidence which suggests that a majority of students exhibit some level of indecision during their postsecondary experience. Gordon (2007) notes that even students who are seemingly decided, in that they have officially declared a major, may still have uncertainty about their major or career choice. This is evidenced by the large number of students who change their major at least once as they progress through college. Cuseo (2005, p. 6) summarizes several findings that illustrate the prevalence of major and career indecision among undergraduate students:

1. Three of every four students are uncertain or tentative about their career choice at college entry (Titley & Titley, 1980; Frost, 1991)
2. Among first-year students who enter college with a major in mind, less than 10% feel they know “a great deal about their intended major” (Lemoine, cited in Erickson & Summers, 1991)
3. Uncertainty among new students frequently increases rather than decreases during their first two years of college (Tinto, 1993)
4. Over two-thirds of entering students change their major during their first year (Kramer, Higley, & Olsen, 1993)
5. Between 50-75% of all students who enter college with a declared major change their mind at least once before they graduate (Foote, 1980; Gordon, 1984; Noel, 1985)
6. Only one senior out of three will major in the same field they preferred as a freshman (Willingham, 1985).

**Undecided Students’ Motivation for Coursework**

Value seems to be a prominent construct in discussions of students’ motivation for general education and major coursework. Students have higher value for coursework that is perceived as relevant to their future careers (i.e., has utility value), with coursework in the major being a prime example of highly valued coursework. Students who have decided on a major are arguably in a better position than undecided students to judge their coursework’s career relevance. However, a common perception is that students do not consider general education courses to be relevant, with the exception of general education coursework in their major (e.g., a psychology major sees relevance in a general education psychology course). On the other hand, general education may be seen as useful to undecided students who utilize the curriculum's diversity to help them select a major. Therefore, undecided students are expected to have higher utility value for general education than decided students, not only because these students are less focused on specialization, but also because general education can assist them in achieving the short-term goal of selecting a major.

Because utility value tends to be moderately to highly correlated with intrinsic and attainment value, undecided students may have higher *general* value for general education than decided students. One explanation in support of their higher intrinsic value is that undecided students may have a more diverse range of interests (Cuseo, 2005) that are represented in the general education curriculum. Because undecided students tend to have less career-focused reasons for attending college than decided students (Baird, 1967, in Gordon, 2007), they may also see greater importance (attainment value) in general education than declared students.

## Establishing a Measure of Motivation for Coursework: The EVaCS

There are many variables that could potentially influence students' motivation for coursework at different points in time—e.g., proportion of credits taken in general education courses, attitudes toward general education, and major/career decidedness. In order to examine whether and how such variables influence motivation across the academic career, it is imperative to use a trustworthy measurement of student motivation. The Expectancy, Value, and Cost Scale (EVaCS; Barron & Hulleman, 2010) measures students' motivation for all courses taken in a particular semester. At the university where the present research was conducted, the EVaCS is administered to students as part of university-wide assessment activities at two time points: upon entering the university (as *incoming* students) and after attaining 45-70 credits (as *mid-career* students).

The EVaCS is based on expectancy-value theory (Eccles et al., 1983) and thus incorporates the value construct that is central to motivation for general education and major coursework. The EVaCS also incorporates the expectancy and cost constructs. Theory and prior research have highlighted different types of expectancy, value, and cost (e.g., Eccles et al., 1983; Flake, 2012). In order to assess motivation in as few items as possible, the EVaCS was designed to measure these constructs at a general level. For instance, rather than including several items to measure attainment, intrinsic, and utility value separately, the EVaCS measures *general* value using a small number of items that span attainment, intrinsic, and utility value. The EVaCS is intended to produce three scale scores (Expectancy, Value, and Cost), which together summarize students' motivation for their courses in a semester. The EVaCS' theoretical grounding is necessary, but not sufficient, for it to be regarded as a good measure of students' motivation. Evidence of the scale's structural integrity is needed as well. Although some research has been conducted on the structure of similar motivation instruments (Flake et al., 2011; Getty et al., 2013; Kosovich, 2013), additional research on the EVaCS' structure is needed if it will be used to make valid inferences about motivation for coursework across the academic career.

**Benson's Framework for Construct Validation**

Benson (1998) provides a three stage framework for construct validation describing what evidence is needed to support valid inferences. The process of creating the EVaCS is represented by the first stage of Benson's (1998) framework—the substantive stage—which involves defining theoretical constructs (e.g., expectancy, value, and cost) and writing items that align to theory. The second stage—the structural stage—involves examining how the relationships among the items align with the instrument's proposed factor structure. A correlated three-factor structure has been established through research on similar measures (e.g., Flake et al., 2011), but the EVaCS' factor structure has not yet been studied. The structural stage is also concerned with whether an instrument functions equivalently in different situations, such as across different time points or between different samples. In this instance, the longitudinal measurement invariance of the EVaCS has not yet been established. The third stage of Benson's (1998) framework—the external stage—involves examining how scores from the instrument relate to other variables. In particular, this stage is concerned with whether the instrument relates to other variables in theoretically predicted ways. The question of how motivation changes over time as the proportion of general education courses changes is an example of an external stage inquiry. However, external stage research questions rely on having substantive (stage one) and structural (stage two) evidence for construct validation.

The present research investigated two main structural stage questions: whether the hypothesized three-factor model fits the EVaCS data better than alternative models, and whether the EVaCS is an invariant measure of motivation for coursework over time. In addition, if the instrument exhibits longitudinal invariance, then it is possible to test whether there are latent mean differences in Expectancy, Value, and Cost for incoming and mid-career students. Establishing a well-fitting, parsimonious, invariant measurement model that is also theoretically supported would enable researchers and practitioners to use the EVaCS to address a variety of external stage research questions regarding students' motivation for semester coursework.

## Conclusion

Value, an important component of students' academic motivation, is likely to differ across the college curriculum. Although students tend to report high value for their major coursework, there is evidence that students' value for general education is lacking, and the reason for this seems due to the prominent focus among students on career preparation. However, students who are undecided in terms of their major or future career may not experience as large of a motivational discrepancy for major and general education coursework. Because students' coursework differs across the academic career, and motivation for different types of coursework (general education vs. major) differs as well, it is reasonable to investigate whether motivation changes across the academic career. To do so, a measure of motivation—in particular, one that captures value—that is theoretically grounded and functions equivalently across time is needed. In other words, the measure must exhibit longitudinal measurement invariance (Vandenberg & Lance, 2000). Chapter 3 outlines how these measurement issues were addressed through factor analytic and invariance studies of an instrument for motivation in coursework, the EVaCS. Specifically, the following three research questions were addressed:

1) Does the hypothesized three-factor model fit the EVaCS data better than the alternative models tested?
2) Does the EVaCS exhibit longitudinal measurement invariance for incoming and mid-career students?

3) Do the latent means of motivation for coursework differ for incoming and mid-career students?

# CHAPTER 3.

# METHOD

This chapter describes the procedures and analyses that were used to address the research questions.  Data collection, samples, and the measurement instrument are described first, followed by an overview of the planned analyses for two studies.  Study 1 examined the model-data fit of four models through confirmatory factor analysis (research question 1).  Study 2 examined whether the EVaCS exhibited longitudinal measurement invariance across two time points (research question 2).  If the EVaCS exhibits measurement invariance, then changes in students' motivation for coursework over time can be examined at the latent level (research question 3).

## Data Collection

Data were collected from undergraduate students at a midsized public university at two time points.  Time point 1 (T1) occurred in late August 2011 when all incoming freshmen were required to participate in the university's Fall Assessment Day.  Fall Assessment Day is held annually on the Friday immediately prior to the start of fall semester classes.  On Assessment Day, students completed a variety of instruments pertaining to general education knowledge, beliefs, attitudes, and behaviors.  Each student was assigned to one of approximately 30 proctored testing rooms—each of which had a specific instrument sequence—based on the last two digits of their student ID number.[7]  Because ID numbers had been assigned by the university essentially at random, this procedure ensured that a random sample of students completed each instrument.  Within each testing room, students completed anywhere from four to six different

[7] Approximately 10 percent of students completed their assessments via computers in campus labs; the rest completed paper-and-pencil assessments in a classroom setting.  Data were treated the same regardless of completion mode.

instruments over a three-hour period. Most instruments were administered in a subset of testing rooms (i.e., to only a sample of students); however, certain instruments were administered in all rooms (i.e., completed by all students) on Assessment Day. The first instrument administered in each room was a general education knowledge test (e.g., quantitative and scientific reasoning), and the second was the Attitudes toward Learning instrument (ATL). The Expectancy, Value, and Cost Scale (EVaCS) was the first scale to appear on the ATL; thus, it was the first non-cognitive scale completed by all incoming freshmen on Assessment Day.

Time point 2 (T2) occurred in mid-February 2013 when all students with 45-70 earned credits were required to participate in the university's annual Spring Assessment Day. The procedures for Spring Assessment Day were basically identical to those for Fall Assessment Day described above. For example, the EVaCS was the first scale to appear on the ATL, which was administered as the second instrument in each testing room (after a general education test, which was the first). Student ID numbers were assigned to testing rooms—and instruments—in Spring 2013 exactly as they had been in Fall 2011. This assessment design facilitated longitudinal comparisons because a majority of students at the university completed the same instruments first as incoming freshmen and again as mid-career students (either sophomores or juniors). Many students who completed assessments at T1 in Fall 2011 also did so at T2. Those who were not assessed at T2 may have withdrawn from the university or had not yet earned enough credits to be eligible for Spring Assessment Day in February 2013.[8] Similarly,

[8] In rare instances, a student who was assessed at T1 may have earned too *many* credits (over 70) to be eligible for Spring Assessment Day at T2. In addition, students who were assessed at T1 may have missed Spring Assessment Day and later attended a mandatory makeup testing session. Data from makeup sessions were not analyzed in the present research.

students who were assessed at T2 but not T1 may have transferred into the university or participated in an earlier Fall Assessment Day (e.g., August 2010) but did not meet the 45-70 credit eligibility requirement for Spring Assessment Day until February 2013.

**Samples**

Because the planned data analyses require multiple large samples, the data were preliminarily screened to verify that sample sizes would be sufficiently large to address the research questions. In total, 3,832 incoming students completed the EVaCS at T1, and 3,324 mid-career students completed the EVaCS at T2. Of the incoming students at T1, 3,749 (98%) had complete data for all EVaCS items. Of the mid-career students at T2, 3,290 (99%) had complete data for all EVaCS items. A matched sample of 2,312 students had complete data at both T1 and T2. There were 2,415 students with complete data at only one time point (i.e., unmatched): 1,437 students had complete data at T1 but not T2, and 978 students had complete data at T2 but not T1.

Five independent samples were formed to conduct the analyses in Studies 1 and 2 (see Table 2). Because the most complex CFA model tested (the bifactor model) has 39 estimated parameters, a minimum sample size of 780 was needed to meet the 20:1 cases-to-parameter guideline for model-fitting (Kline, 2013). First, a randomly selected portion of the matched sample ($N$ = 951) was reserved for Study 2 to address the research question of longitudinal measurement invariance. From the remaining cases, four independent samples of 944 each were formed for Study 1 (model-data fit). Each of these samples was comprised of 50 percent incoming and 50 percent mid-career students. Excluding the cases reserved for Study 2, every student with complete data for at least one time point was assigned to one of the four model-fitting samples. This included a portion of cases from the initial matched sample that were randomly chosen to supply data at *either* T1 ($N$ = 451) *or* T2 ($N$ = 910) in order to achieve time point balance within each sample. After determining the time point for which each case would supply data, the 1,888 T1 cases and 1,888 T2 cases were randomly assigned to one of four samples ($N$ = 944 each).

Samples 1 and 2 were used to test and cross-validate the proposed models, described below. Cross-validation is necessary to avoid capitalizing on idiosyncrasies from any particular sample (MacCallum, Roznowski, & Necowitz, 1992), and it is particularly important for this research because extensive psychometric work has not yet been conducted on the EVaCS. Results from Samples 1 and 2 indicated that model modifications were necessary, so two additional samples (Samples 3 and 4) were used to test and cross-validate the fit of modified models. As previously stated, longitudinal measurement invariance was examined in Study 2 using the sample of 951 matched cases.

## Measure

### Expectancy, Value, and Cost Scale (EVaCS)

The 16 EVaCS items (Appendix A) are based on expectancy-value theory and were written to assess general expectancy, value, and cost across all of the courses students take in a semester. For example, the value items are intended to represent the breadth of the value construct rather than to serve as indicators of specific types of value (e.g., attainment value). Unlike many other motivation scales based on expectancy-value theory, the EVaCS includes several items intended to measure cost, which the scale developers argue to be theoretically distinct from value (Barron & Hulleman, 2010; Flake et al., 2011). Students responded to each item on a 1 (*completely disagree*) to 8 (*completely agree*) scale. No items were reverse coded. Therefore, higher scores represented greater expectancy, more value, and higher cost. Thus, a student who has optimal motivation should select high responses to expectancy and value items (which enhance motivation) and low responses to cost items (which inhibit motivation).

## Study 1: Assessing Model-Data Fit Using Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is a modeling technique used when one or more unobservable, latent traits are believed to underlie (i.e., be influencing) item responses (Kline, 2011). In testing model-data fit, CFA examines the theoretical relationships among observed item responses and latent traits according to specified models. The EVaCS is developed from expectancy-value theory, so students' responses to the scale's items are thought to be driven by their achievement motivation—specifically, students' expectancy, value, and cost (latent traits) for a semester's coursework. For example, a student who expects to do well in his courses is high on the latent expectancy trait and should therefore strongly endorse the items intended to measure expectancy. A student who experiences low cost related to succeeding in his coursework should select lower responses to the items intended to measure cost. Such theoretical relationships are specified a priori and tested via CFA. In addition to testing a theoretical model, one or more alternative models may be tested. If a theoretical model yields better fit to the data than alternative models, this provides greater support for the theoretical model.

The purpose of Study 1 was to test and compare the fit of several models, which are illustrated in Figures 1-4 and described in detail below: (a) a unidimensional model,

(b) a correlated two-factor model, (c) a bifactor model, and (d) a correlated three-factor model.

**Unidimensional Model**

The unidimensional, or one-factor, model treats all 16 EVaCS items as indicators of a single latent motivation trait (see Figure 1). A well-fitting unidimensional model would mean that the EVaCS items do not represent distinct latent factors. Therefore, when scoring the EVaCS using the unidimensional model, each student would receive one total Motivation scale score based on all 16 items. Given that previous research (e.g., Eccles & Wigfield, 1995; Flake et al., 2011; Kosovich, 2013; Kosovich, Hulleman, & Barron, 2013) has found similar items separate into at least two factors (expectancy and value), this model was not hypothesized to fit well. However, if the unidimensional model does fit the data well, this would indicate that the items do not represent three distinct factors as the scale developers intended, and thus, item revisions would be necessary to better align the scale with expectancy-value theory. The unidimensional model has 104 degrees of freedom; 32 parameters (16 factor loadings, 16 error variances) are estimated from 136 observations.

**Correlated Two-Factor Model**

This model more closely represents classic expectancy-value theory in that Expectancy and Value are two distinct, although related, latent factors (see Figure 2).[9] Four items load onto the Expectancy factor. The other twelve items load onto the Value factor. The Expectancy and Value factors are correlated. When scoring the EVaCS

---

[9] To simplify the description of models and reporting of results, factors will be referred to by the theoretical constructs the items are intended to measure—i.e., Expectancy, Value, and Cost. More validity evidence (beyond structural evidence) is needed to support the claim that these are indeed the constructs responsible for systematic variance in EVaCS item responses (Benson, 1998).

using the two-factor model, a student would receive two scale scores: an Expectancy score based on 4 items and a Value score based on the other 12 items. This model does not distinguish among cost and other types of value (attainment, intrinsic, utility) specified in the literature. Therefore, a well-fitting model would suggest that cost is a subtype of value, as proposed by Eccles and colleagues (Wigfield & Eccles, 2000). However, because 6 of the 12 value items are intended to measure cost, which generally does not have high correlations with the other value subtypes (Conley, 2012; Trautwein et al., 2012), there may be substantial shared error in the cost items that is not captured by the Value factor. For this reason, the correlated two-factor model may not fit as well as a model that distinguishes cost from value. However, if the correlated two-factor model does fit the data well, this would indicate that the items represent two factors instead of the three distinct factors as the scale developers intended. Further work would then be needed to determine whether cost separates from more differentiated measures of value (i.e., attainment, intrinsic, utility) as opposed to the EVaCS' general measure of value. The correlated two-factor model has 103 degrees of freedom; 33 parameters (16 factor loadings, 16 error variances, 1 factor correlation) are estimated from 136 observations.

**Bifactor Model**

The most complex model tested was the bifactor model. This model represents classic expectancy-value theory in that cost is considered a value subtype (Wigfield & Eccles, 2000; see Figure 3). Four items load onto the Expectancy factor. The other twelve items all load onto a general Value factor. The Expectancy and Value factors are correlated. In addition, a proportion of six value items' variance that is *not* due to the general factor is believed to be due to another, unrelated specific factor—Cost.[10] Because these six items load on two factors, the bifactor model presents practical challenges for scoring (by summing item responses) and interpretation that do not exist with the other models. A student would receive an Expectancy score based on 4 items; however, a Value score would be based on the other 12 items, 6 of which would include systematic variance from the specific Cost factor. In other words, the Value score would knowingly include item variance from Cost which, in the model, is specified as *unrelated* to Value. This would not be terribly concerning if the double-loading items (i.e., the six items with loadings on Value and Cost) load strongly onto the Value factor. If the double-loading items have strong factor loadings (e.g., .70 or greater) on Value, this would indicate that cost can be treated as a subtype of value. Weaker factor loadings on Cost would represent that these items share variance with each other that they do not share with the other six items that load onto the Value factor. The six double-loading items could therefore still be justified as measuring general value, even though something other than value makes them function somewhat distinctly as cost items. However, if these six items have stronger loadings on the Cost factor than they have on the Value factor, this would indicate that most of their variance is explained by something unrelated to value. Therefore, these items would be primarily measuring something distinct from value—cost—and thus to include them when computing a Value scale score would be erroneous. The bifactor model was expected to fit better than the previous two models

[10] It may seem strange to consider cost as a subtype of value when, in the bifactor model, the Cost factor is specified as *unrelated* to the Value factor. In this model, the Cost factor represents the variance shared among items 11-16 that is unrelated to the variance shared among items 5-16 (*all* of the value items). In other words, the Cost factor represents what makes items 11-16 function similarly after accounting for what makes them function similarly to items 5-10. In order for cost to be regarded as a subtype of value, the Value factor should explain more of the variance in items 11-16 than the Cost factor explains.

because (1) unlike the unidimensional model, it distinguishes between expectancy and value, and (2) unlike the two-factor model, it attempts to model systematic variance in the cost items. The bifactor model has 97 degrees of freedom; 39 parameters (22 factor loadings, 16 error variances, 1 factor correlation) are estimated from 136 observations.

**Correlated Three-Factor Model**

This is the model that the EVaCS items are intended to represent, in which Expectancy, Value, and Cost are distinct, although related, latent factors (Barron & Hulleman, 2010; see Figure 4). Four items load onto the Expectancy factor, six items load onto the Value factor, and the remaining six items load onto the Cost factor. These three factors are correlated. Using the three-factor model, a student would receive three scale scores: an Expectancy score based on 4 items, a Value score based on 6 items, and a Cost score based on 6 items. This is the only model of the four tested in which cost is modeled separately from value and thus provides a test of the theoretical value-cost relationship. If the correlated three-factor model were the best-fitting model among the four tested, it would indicate that cost is most appropriately treated as a distinct factor rather than a subtype of value. This model was expected to fit the data well because (1) it most closely represents the model the EVaCS was designed to measure, and (2) previous research on similar scales has found that cost is best modeled as a separate factor from value (Flake et al., 2011; Kosovich, 2013). The correlated three-factor model has 101 degrees of freedom; 35 parameters (16 factor loadings, 16 error variances, 3 factor correlations) are estimated from 136 observations.

### Determining the Best-Fitting Model

The procedures for determining which of the four models provided the best model-data fit relied on a variety of fit indices: RMSEA, SRMR, CFI, $\chi^2$ difference tests (for nested models) or AIC (for non-nested models), and correlation residuals. Fit indices were interpreted relative to commonly used guidelines in the literature (e.g., Hu & Bentler, 1998; 1999; Kline, 2011). These fit indices are described in detail in Chapter 4.

## Study 2: Testing for Measurement Invariance and Latent Mean Differences

Prior to examining change in the motivation of incoming and mid-career students, it is important to assess whether the structure of the latent traits and parameters of a measurement model are equivalent over time (Byrne, Shavelson, & Muthén, 1989). That is, longitudinal measurement invariance must first be established to ensure that respondents continue conceptualizing the construct(s) in the same way and that the instrument functions consistently when administered at multiple time points. If an instrument lacks longitudinal invariance, then one cannot be certain that any observed change in scores is due to a true change at the latent level. Furthermore, without longitudinal invariance, one cannot be certain that an absence of observed change signifies no latent change. Longitudinal measurement invariance was evaluated by first testing configural invariance, then metric invariance, then scalar invariance.

### Configural Invariance

Configural invariance—a prerequisite for metric and scalar invariance—means that the same measurement model provides adequate fit in the same sample over time. For example, if the EVaCS demonstrates configural invariance, this means that a model which fit the data from incoming students also fits the data when the same students are assessed later as mid-career students. Using the correlated three-factor model to illustrate, configural invariance would indicate that the items that serve as indicators for Expectancy at T1 are also indicators of Expectancy at T2 (likewise for Value and Cost indicators at T1 and T2). As a preliminary step toward testing configural invariance, the best-fitting model from Study 1 can be fit separately to data from T1 and T2. If the model shows adequate fit at both time points according to various local and global fit indices (e.g., RMSEA, CFI, correlation residuals), then the model is fit to T1 and T2 simultaneously.[11] If this model fits, then the EVaCS demonstrates configural invariance.

**Metric Invariance**

If configural invariance is established, the next step is to test metric invariance, which means that factor loadings are equivalent over time (i.e., indicators have an equivalent relationship with their respective latent factor at both time points). If the EVaCS demonstrates metric invariance, it would mean that an item's unstandardized pattern coefficient in the model at T1 is essentially equal to the corresponding coefficient at T2. For example,

$$\lambda_{\text{item1 at T1}} = \lambda_{\text{item1 at T2}},$$

$$\lambda_{\text{item2 at T1}} = \lambda_{\text{item2 at T2}},$$

$$\ldots$$

$$\lambda_{\text{item16 at T1}} = \lambda_{\text{item16 at T2}}$$

[11] When fitting the model to T1 and T2 data simultaneously, item residuals are correlated across time (e.g., $\varepsilon_{\text{item1 at T1}}$ is allowed to correlate with $\varepsilon_{\text{item1 at T2}}$). These *autocorrelations* reflect shared variance among the same item across testing occasions. In addition, the factor(s) at each time point are allowed to correlate. In the longitudinal bifactor model, the specific factor (Cost) is allowed to correlate over time, but not with the substantive factors (i.e., the Cost factor does not correlate with Expectancy or Value factors at either time point).

Metric invariance would indicate that all items are measuring the construct in the same way across time; i.e., items are interpreted the same way by incoming and mid-career students. Another way to understand metric invariance is that items have the same salience to their respective factors at both time points. To test for metric invariance, unstandardized pattern coefficients are constrained to be equal across T1 and T2. If this results in adequate model-data fit (i.e., if the metric model fits in an absolute sense and does not fit significantly worse than the configural model), then the EVaCS demonstrates metric invariance. Fit of the metric model is compared to that of the configural model using a $\chi^2$ difference test and examining change in CFI. In addition, residuals between the observed and reproduced correlation matrix are examined to determine whether any pattern coefficients should not be constrained to be invariant across time (i.e., whether there is partial metric invariance).

**Scalar Invariance**

When metric invariance is established, it becomes possible to test scalar invariance. Scalar invariance indicates that in addition to unstandardized pattern coefficients being equivalent for items across time (metric invariance), item intercepts are set to be equal for items across time. For example,

$$\tau_{\text{item1 at T1}} = \tau_{\text{item1 at T2}},$$

$$\tau_{\text{item2 at T1}} = \tau_{\text{item2 at T2}},$$

$$\ldots$$

$$\tau_{\text{item16 at T1}} = \tau_{\text{item16 at T2}}$$

With equal intercepts, a student with the same level of a latent trait (e.g., expectancy) at multiple time points would also have the same observed item scores across time points.

Therefore, any *differences* in observed means would be due to true differences at the latent level. For example, if the EVaCS exhibits scalar invariance, then a student whose score on Expectancy item 1 increases from T1 to T2 can be said to have truly increased in his expectancy for semester coursework between the two time points. To test for scalar invariance, item intercepts are constrained to be equal across T1 and T2. If this results in adequate model-data fit (i.e., if the scalar model fits in an absolute sense and does not fit significantly worse than the metric model), then the EVaCS demonstrates scalar invariance. Fit of the scalar model is compared to that of the metric model using a $\chi^2$ difference test and examining change in CFI. In addition, residuals between the observed and reproduced correlation matrix are examined to determine whether any item intercepts should not be constrained to be invariant across time (i.e., whether there is partial scalar invariance).

**Latent Mean Differences**

If configural, metric, and either full or partial scalar invariance are established for the EVaCS, then latent mean scores for incoming and mid-career students can be compared to determine whether students change in their levels of motivation for semester coursework. For example, full scalar invariance using the correlated three-factor model would imply that latent means of expectancy, value, and cost can be compared across time. Partial scalar invariance would imply that latent means can be compared across time for any latent trait with at least three scalar-invariant items (Thompson & Green, 2006), where only the scalar-invariant items contribute to the latent means. When scalar or partial scalar invariance are established for a latent trait, latent mean differences can be examined for statistical significance (using a latent *t*-test) and practical significance (using a latent Cohen's *d*). When metric invariance is established, the stability of latent traits can be compared through bivariate correlations, which represent the stability of rank-ordered latent traits over time.

## Conclusion

The purpose of this research was to evaluate the psychometric structure and invariance of a scale intended to measure students' motivation for their semester coursework—the EVaCS. First, the model-data fit of four models was tested through CFA using multiple samples of incoming and mid-career students. The best-fitting model was then used to examine longitudinal measurement invariance (configural, metric, and scalar) of the EVaCS for students at two time points: (1) as incoming freshmen, and (2) after earning 45-70 credits, or as "mid-career" students. With measurement invariance established, latent mean change in motivation for semester coursework can be examined for statistical and practical significance using a latent *t*-test and Cohen's *d*, respectively.

# CHAPTER 4.

# RESULTS

## Study 1

The purpose of Study 1 was to determine which of four a priori models provided the best, parsimonious fit to the EVaCS data by testing the models through confirmatory factor analysis (CFA). Four independent samples, each comprised of equal parts incoming and mid-career students, were used to test and cross-validate model fit to arrive at the best-fitting model, which would later be used to examine measurement invariance in Study 2.

### Data Screening

Only cases with complete data at time point 1 (T1) and/or time point 2 (T2) were used in the analyses. Data were screened for outliers, multicollinearity, and normality prior to conducting the analyses. In addition, preliminary reliability analysis (Cronbach's α) on the 16 EVaCS items indicated that item 13 ("I think there will be other things I'd rather do with my time than just focusing on my classes") negatively impacted the reliability of the intended Cost scale in all four samples in Study 1. Therefore, item 13 was removed from the models, leaving 15 EVaCS items to be included in the CFA models.

**Outliers.** Data cases are considered outliers if an individual item response (univariate) or overall response pattern (multivariate) is considered aberrant, and such cases should be removed prior to analyses. All item responses were within the EVaCS' 1-8 response scale range, as out-of-range responses had previously been converted to missing data and removed through listwise deletion.

A case is deemed a multivariate outlier when the pattern of item responses differs substantially from the norm, such as when a respondent engages in random responding or response sets across items (e.g., ‘1’, ‘2’, ‘3’, ‘4’, ‘5’, ‘4’, ‘3’, etc.). All cases in Samples 1 through 4 were simultaneously screened for multivariate outliers. Prior to the removal of multivariate outliers, each sample had 944 cases. A total of 94 cases (2.5%) were identified as multivariate outliers based on having Mahalanobis distances greater than 51.25 ($p < .05$) and were removed from the data set. The remaining sample sizes for Samples 1 through 4 were 921, 923, 916, and 922, respectively. Data screening and analyses proceeded with these reduced sample sizes. Descriptive statistics and demographic information for each sample are shown in Table 3.

**Multicollinearity.** If an item is found to be multicollinear, this means that it is essentially redundant with another item(s) and thus does not provide unique information. One or more redundant items from a set deemed multicollinear can be considered for removal from the analyses. Multicollinearity was assessed by first examining the bivariate correlations among items in each sample. If two items are highly correlated ($|r| > .90$; Tabachnick & Fidell, 2005), then one item from the pair is essentially redundant. Tables B1-B4 show the bivariate correlations for each sample in Study 1. The largest observed correlations were between items 6 (“I think my classes will be worthwhile”) and 7 (“I think my classes will be useful”), ranging from $r = .73$ to .76 across samples, and between items 9 (“I see a purpose for taking my classes this semester”) and 10 (“I see why my classes are important to take”), ranging from $r = .74$ to .79 across samples. Because no correlations exceeded .90, bivariate multicollinearity was not a problem.

Multivariate multicollinearity, which reflects whether items on a scale provide unique information, was assessed through tolerance values. A low tolerance value (<.10) indicates that most of an item's variance can be explained by other item(s) on the scale, so that item provides little unique information. Conversely, items with high tolerance values cannot be explained primarily by other items, and therefore contribute unique information to a scale. Tolerance values for all EVaCS items were greater than .10, so multivariate multicollinearity was not a problem.

**Normality.** The maximum likelihood estimation method used in the analyses assumes that observed data have a multivariate normal distribution. If data are not normally distributed, then the resulting standard errors, $\chi^2$ values, and fit indices will be biased (West, Finch & Curran, 1995). Univariate non-normality was identified by skewness statistics with absolute values of 2 or greater and/or kurtosis statistics with absolute values of 7 or greater (West et al., 1995). As shown in Tables B1-B4, the largest skewness and kurtosis statistics across samples were 1.05 and 1.46, respectively. Items 1 through 10 all had slight negative skews, and the remaining items had slight positive skews; however, none of the skewness or kurtosis statistics were large enough to indicate univariate non-normality.

Multivariate normality was assessed in each of the four samples through Mardia's coefficient for multivariate kurtosis (DeCarlo, 1997). Because the coefficients, which ranged from 37.12 to 46.79 across samples, all exceeded a value of 3 (Bentler & Wu, 2003), the Satorra-Bentler adjustment was applied to the $\chi^2$ values, $\chi^2$ difference tests, standard errors, and fit indices produced through maximum likelihood estimation (Satorra & Bentler, 1994). By doing so, the Satorra-Bentler adjustment is said to produce less biased estimates of model fit and standard errors for parameter estimates (Finney & DiStefano, 2006).

**Descriptive Statistics**

Item-level descriptive statistics for each sample are shown in Tables B1-B4; they were similar across samples. Items 1 through 10 (hypothesized to measure expectancy and value) had rather high means on the 8-point scale. Sample means for the intended expectancy items ranged from 6.41 on item 4 ("I know I can understand the material in my classes") to 6.77 on item 1 ("I expect to do well in my classes"). Sample means for the intended value items ranged from 6.32 on item 8 ("I think my classes will be interesting") to 6.75 on item 5 ("I value the classes I am taking this semester"). Items 11, 12, 14, 15, and 16 had lower means, believed to reflect students' lower levels of cost. Sample means for these items ranged from 2.56 on item 15 ("Doing well in my classes may not be worth all the things I have to give up") to 3.81 on item 12 ("Because of the other things that I want to do in college, I don't think I'll have as much time to put into my classes as I'd like"). Across all samples, the range of selected responses was 1-8 for items 9 through 16, 2-8 for items 2 through 8, and 3-8 for item 1. Therefore, students utilized the full response scale—or nearly did so—for all of the EVaCS items.

## Confirmatory Factor Analysis

Four a priori models were each fit to data from four independent samples: (a) a unidimensional model, (b) a correlated two-factor model, (c) a bifactor model, and (d) a correlated three-factor model. Samples 1 and 2 were used to fit the 15-item models (without item 13). Samples 3 and 4 were used to test modifications to these models—specifically, removing any item(s) that contributed to model misfit in Samples 1 and 2.

All models were tested in LISREL 8.80 (Jöreskog & Sörbom, 2006) using maximum likelihood estimation with Satorra-Bentler scaled $\chi^2$ and robust standard errors to adjust for multivariate non-normality. The metric of latent factors was set equal to 1 by default in LISREL.

**Fit Indices**

Several different fit indices were used to evaluate model fit. These included indices of global fit (absolute and incremental) and local misfit. The root mean square error of approximation (RMSEA) and standardized root mean squared residual (SRMR) are two absolute fit indices that are sensitive to model misspecification (Hu & Bentler, 1998). Whereas RMSEA is more sensitive to complex model misspecification (i.e., constraining factor loadings to zero), SRMR is more sensitive to simple model misspecification (i.e., constraining factor correlations to zero). The comparative fit index (CFI) is an incremental fit index that compares the fit of a theorized model to the fit of a model in which all observed variables are unrelated. Yu and Muthén (2002) provide general guidelines for interpreting fit indices when using the Satorra-Bentler adjustment: desirable values are RMSEA $\leq .05$, SRMR $\leq .07$, and CFI $\geq .96$. It is important to note that these values are not intended to be absolute, and it is recommended that multiple fit indices be considered when evaluating model fit.

When two nested models both demonstrated acceptable fit according to RMSEA, SRMR, and/or CFI, $\chi^2$ difference tests ($\Delta\chi^2$) were used to directly compare the fit of these models. Unlike a $\chi^2$ significance test, which indicates whether a model is able to provide exact fit to observed data (i.e., whether the model-implied covariance matrix differs significantly from the observed covariance matrix), the $\chi^2$ difference test can be used to tell whether two nested models differ significantly in their ability to reproduce the same covariance matrix. So, in the event that nested models have comparable global fit indices, a $\chi^2$ difference test can be conducted to determine whether the more complex or constrained model (i.e., the model with fewer degrees of freedom) fits significantly worse than the less complex model. The unidimensional model is nested within each of the other three models. The correlated two-factor model is nested within the bifactor and correlated three-factor models. The $\chi^2$ difference test only applies to nested models (Kline, 2011), so the Akaike Information Criterion (AIC) was used to compare the fit of non-nested models—specifically, the bifactor and correlated three-factor models. The AIC adjusts $\chi^2$ values according to model parsimony by penalizing more complex models, particularly with small sample sizes (Mulaik, 2009). A model with a lower AIC should generally be favored over a model with a higher AIC because it is more likely to replicate in other samples due to its greater parsimony.

Global fit indices such as RMSEA, SRMR, and CFI can indicate that a particular model fits the data well overall, but there may be specific areas of the model which do not produce good model-data fit. Therefore, local misfit was assessed in globally-fitting models. Correlation residuals represent the difference between the correlation matrix from the observed data and the correlation matrix as reproduced (or implied) by the model. Residuals of |.10| or larger indicate areas of local misfit where a model is not able to reproduce item correlations well (Kline, 2011). Items associated with large correlation residuals contribute to model misfit and may be candidates for removal in modified models.

**Testing and Cross-Validating 15-Item Model Fit in Samples 1 and 2**

Global fit indices from the models tested in Samples 1 and 2 are presented in Table 4.

**Unidimensional model.** The unidimensional model, which treated all EVaCS items as indicators of a single latent factor, did not achieve adequate global fit in either sample, with all indices failing to meet their respective cutoffs (RMSEA .19-.20, SRMR .13-.14, CFI .76-.79). Therefore, local misfit of this model was not examined.

**Correlated two-factor model.** The two-factor model, in which cost was *not* modeled as a subtype of value, did not achieve adequate global fit in either sample, with all indices failing to meet their respective cutoffs (RMSEA .15-.16, SRMR .12-.13, CFI .84-.87). Therefore, local misfit of this model was not examined.

**Bifactor model.** This model, which modeled cost as a subtype of value, met the cutoffs for SRMR (.06-.07) and CFI (.96-.97) in both samples, and nearly did for RMSEA (.08). AIC values for the bifactor model were 664.92 in Sample 1 and 622.38 in Sample 2. Recall that AIC values can be compared across non-nested models (i.e., the bifactor and three-factor models) to evaluate fit adjusting for model parsimony. Standardized pattern coefficients—also referred to as “factor loadings”—and error variances are presented in Figures 5 and 6 for Samples 1 and 2, respectively. The correlation between the Expectancy and Value latent factors was .62 in Sample 1 and .67 in Sample 2. Items 1 through 10 loaded strongly onto their respective factors (.71 to .86), and the latent factors accounted for the majority of the variance in each of these 10 items ($R^2$ = .51-.74). The five cost items had moderate loadings on the Cost factor (.48 to .73) but weaker, negative loadings on the Value factor (-.24 to -.40). The proportion of variance ($R^2$) accounted for in these items ranged from .29 to .68.

Local misfit of the bifactor model was assessed via correlation residuals with absolute values greater than .10 (see Table D1). Of the 105 correlations that were reproduced by the bifactor model, 14 correlations (13%) had large residuals in Sample 1, and 13 correlations (12%) had large residuals in Sample 2. Most of the large residuals were between expectancy and cost items, including the largest residual of -.22 between item 1 and item 14. These residuals were negative, meaning that the bifactor model overestimated the relationship between many expectancy and cost items.

**Correlated three-factor model.** This model tested Cost as a distinct factor from Value. Like the bifactor model, this model met the cutoffs for SRMR (.04) and CFI (.97) in both samples, and nearly did for RMSEA (.07-.08). AIC values for the three-factor model were 609.88 in Sample 1 and 565.82 in Sample 2; both were lower (i.e., more favorable) than AIC values for the bifactor model. Standardized pattern coefficients and error variances are presented in Figures 7 and 8 for Samples 1 and 2, respectively. The correlation between the Expectancy and Value latent factors was .61 in Sample 1 and .66 in Sample 2. Both factors had negative correlations with Cost (Expectancy: -.50 in Sample 1, -.53 in Sample 2; Value: -.42 in Sample 1, -.45 in Sample 2). The majority of items had factor loadings greater than .70. Two exceptions were item 11, with a loading of .55 in Sample 1 and .58 in Sample 2, and item 12, with a loading of .70 in Sample 1 and .67 in Sample 2. The latent factors accounted for a majority of the variance in most items ($R^2$ = .51-.75), with the exceptions of item 11 ($R^2$ = .30-.33) and item 12 ($R^2$ = .44-.49).

Local misfit of the three-factor model was assessed via correlation residuals (see Table D2). Of the 105 correlations that were reproduced by the three-factor model, 2 correlations (2%) had large residuals in Sample 1, and 1 (1%) residual was large in Sample 2. All of the large residuals were observed among cost items, but none exceeded |.13|. These residuals were positive, meaning that the three-factor model underestimated the relationship between the items.

**Summary of Results from Samples 1 and 2**

Of the 15-item models tested in Samples 1 and 2, both the bifactor and correlated three-factor models demonstrated acceptable global fit. There are five reasons to favor the three-factor model over the bifactor model at this point. First, the models' AIC values, which adjust $\chi^2$ for model parsimony, favored the three-factor model. Second, the three-factor model had fewer large correlation residuals (1-2 > |.10|) compared to the bifactor model (13-14 > |.10|). Third, double-loading items in the bifactor model had weaker loadings on the Value factor (-.24 to -.40) than on the Cost factor (.48 to .73), indicating that these items bore a stronger relationship to Cost—as a distinct factor—than to Value. Fourth, the three-factor model is consistent with prior research from scales with items similar to those on the EVaCS. Fifth, the three-factor model yields scoreable subscales, whereas in the bifactor model, variance from the Value and Cost factors cannot be fully disentangled to produce separate subscales.

Although the three-factor model fit quite well with 15 items, it was clear that item 11 ("I think my classes will require too much time for me to do well") did not have a strong loading on its intended Cost factor. (The fit of item 12 was also potentially problematic, though not as severe as item 11's). It is important that whichever model is championed in Study 1 demonstrates adequate global and local fit before being subjected to invariance testing in Study 2. Therefore, based on the results from Samples 1 and 2, item 11 was removed, and the four CFA models were retested in Samples 3 and 4 to see whether model fit would improve. If model fit continued to show problems even after the removal of item 11, additional model modifications would be considered in Samples 3 and 4.

## Testing and Cross-Validating 14-Item Model Fit in Samples 3 and 4

Due to its poor fit in Samples 1 and 2, item 11 was not included in the CFAs conducted on Samples 3 and 4. Despite their failure to demonstrate adequate global fit in Samples 1 and 2, the unidimensional and correlated two-factor models were retested in Samples 3 and 4 in case their fit would improve after the removal of item 11. Fit indices are presented in Table 4. It is important to note that because the models tested in Samples 1 and 2 differ from those tested in Samples 3 and 4, direct comparison of fit across models is not advised, although fit within each sample can still be interpreted relative to recommended cutoffs.

**Unidimensional model.** This model did not achieve adequate global fit in either sample (RMSEA .18-.20, SRMR .12-.13, CFI .75-.79), so local misfit was not examined.

**Correlated two-factor model.** This model did not achieve adequate global fit in either sample (RMSEA .14-.15, SRMR .11-.12, CFI .86-.90), so local misfit was not examined.

**Bifactor model.** This model met the cutoffs for SRMR (.06) and CFI (.97) in both samples, and nearly did for RMSEA (.07). AIC values for the bifactor model were 451.21 in Sample 3 and 497.41 in Sample 4. Standardized pattern coefficients and error variances are presented in Figures 9 and 10 for Samples 3 and 4, respectively. The correlation between the Expectancy and Value latent factors was .65 in Sample 1 and .66 in Sample 2. Items 1 through 10 loaded strongly onto their respective factors (.72 to .87), and the latent factors accounted for the majority of the variance in each of these 10 items ($R^2$ = .52-.76). The five cost items had moderate loadings on the Cost factor (.58 to .75) but weaker loadings on the Value factor (-.25 to -.38). The proportion of variance ($R^2$) accounted for in the cost items ranged from .39 to .65.

Local misfit was assessed via correlation residuals (see Table D3). Of the 91 correlations that were reproduced by the 14-item bifactor model, 11 correlations (12%) had large residuals in Sample 3, and 10 (11%) had large residuals in Sample 4. As was the case in Samples 1 and 2, most of the large residuals were between expectancy and cost items, including the largest residuals of -.23 between items 1 and 14 and items 3 and 14. These residuals were negative, meaning that the bifactor model overestimated the relationship between many expectancy and cost items.

**Correlated three-factor model.** Like the bifactor model, this model met the cutoffs for SRMR (.03-.04) and CFI (.98) in both samples, and nearly did for RMSEA (.06-.07). AIC values for the three-factor model were 406.86 in Sample 3 and 458.96 in Sample 4, which were lower (more favorable) than the AIC values for the bifactor model. Standardized pattern coefficients and error variances are presented in Figures 11 and 12 for Samples 3 and 4, respectively. The correlation between the Expectancy and Value latent factors was .64 in Sample 3 and .65 in Sample 4. Both factors had negative correlations with Cost (Expectancy: -.50 in Sample 3, -.51 in Sample 4; Value: -.42 in Sample 3, -.46 in Sample 4). All items had factor loadings greater than .70 except item

12, which had a loading of .70 in Sample 3 and .63 in Sample 4. The latent factors accounted for a majority of the variance in most items ($R^2$ = .52-.77), with the exception of item 12 ($R^2$ = .40-.49).

Local misfit was assessed via correlation residuals (see Table D4). Of the 91 correlations that were reproduced by the three-factor model, none (0%) had large residuals in Sample 3, and 3 (3%) had large residuals in Sample 4. None of the large residuals exceeded |.14|.

Parameter and reliability estimates (omega, ω) for the 14-item, three-factor model are shown in Table 5. Across both samples, reliability estimates for the factors were all .80 or higher.

**Summary of Results from Samples 3 and 4**

The bifactor and correlated three-factor models continued to demonstrate acceptable global fit after the removal of item 11. In both samples, AIC values were lower (i.e., more favorable) for the three-factor model than the bifactor model. The three-factor model had fewer large correlation residuals (0-3 > |.10|) compared to the bifactor model (10-11 > |.10|). The bifactor model's double-loading items loaded more strongly onto Cost (.58 to .75) than Value (-.25 to -.38), suggesting a stronger relationship with Cost as a distinct factor from Value. The three-factor model is also consistent with prior research from scales with items similar to those on the EVaCS. Therefore, the three-factor model was selected from the four a priori models to be examined in Study 2.

Although the three-factor model fit quite well with 14 items, the minimal local misfit in Sample 4 appeared to involve item 12—the item which, along with item 11, showed some problems in Samples 1 and 2. At this point, item 12 ("Because of the other things that I want to do in college, I don't think I'll have as much time to put into my classes as I'd like") shows some signs of influencing poor model fit. However, without overwhelming reason to exclude it and refit the data to further-reduced models, item 12 was retained because it adds to the breadth of the cost construct. It should be noted that the Expectancy and Value factors held up well throughout Study 1, so cost items should be considered first if any problems are encountered with invariance testing in Study 2.

# Study 2

Given that a well-fitting model was found in Study 1 (the 14-item correlated three-factor model), the objective of Study 2 was to test whether this model exhibited configural, metric, and scalar invariance, and if so, whether there were latent mean differences between students' motivation for coursework at two time points. Data were fit to a single longitudinal sample of students measured as incoming and as mid-career students ("Matched" sample in Table 2).

## Data Screening

Data screening followed the same procedures for assessing outliers, multicollinearity, and normality as outlined in Study 1. All cases had complete data at time point 1 (T1) and time point 2 (T2).

**Outliers.** All item responses were within the EVaCS' 1-8 response scale range, as out-of-range responses had previously been converted to missing data and removed through listwise deletion. Data from time point 1 and time point 2 were screened separately for multivariate outliers. There were 951 matched cases prior to the removal of multivariate outliers, which were cases with Mahalanobis distances greater than 46.93 ($p < .05$). A total of 16 cases were identified as multivariate outliers at T1, 20 cases at

T2, and 3 cases at both time points. These 39 cases were removed, and data screening and analyses proceeded with a reduced sample size of 912. Descriptive statistics and demographic information for the sample are presented in Table 6.

**Multicollinearity.** Table D1 shows the bivariate correlations for items in Study 2. The largest correlation observed among items at either time point was between items 5 ("I value the classes I am taking this semester") and 6 ("I think my classes will be worthwhile") for time point 2 ($r$ = .80). Overall, correlations among items belonging to the same latent factor were higher at T2 than at T1. Because no correlations exceeded .90, bivariate multicollinearity was not a problem. Tolerance values for all items were all greater than .10, so multivariate multicollinearity was not a problem.

**Normality.** As shown in Table D1, the largest skewness and kurtosis statistics across time points were -.96 and 1.75, respectively, which were not large enough to indicate univariate non-normality. Expectancy and value items all had slight negative skews, while the cost items had slight positive skews. Multivariate normality was assessed at each time point through Mardia's coefficient for multivariate kurtosis (DeCarlo, 1997). The coefficients—27.22 at T1 and 44.81 at T2—exceeded a value of 3 (Bentler & Wu, 2003), so the Satorra-Bentler adjustment was applied to the $\chi^2$ values, $\chi^2$ difference tests, and fit indices (Satorra & Bentler, 1994) as had been done in Study 1.

**Descriptive Statistics**

Item-level descriptive statistics are shown in Table D1. Means for the expectancy items ranged from 6.28 to 6.80 at T1 and from 6.44 to 6.59 at T2. Means for the value items ranged from 6.37 to 6.78 at T1 and from 6.22 to 6.58 at T2. Observed means decreased from T1 to T2 for all expectancy and value items except item 4 ("I know I can understand the material in my classes"). Means for the cost items ranged from 2.27 to 3.50 at T1 and from 2.84 to 3.90 at T2. Observed means increased for all cost items over time. At T1, the range of selected responses was 1-8 for items 12 and 16; 1-7 for item 15; 1-6 for item 14; 2-8 for items 9 and 10; 3-8 for items 4, 5, 7 and 8; and 4-8 for items 1, 2, 3 and 6. So, although students did not utilize the full 1-8 response scale for most items at T1, there was still considerable variability in scores. At T2, the range of selected responses was 1-8 for items 5, 7, 12, 14, 15, and 16; 2-8 for items 1, 6, 8, 9, and 10; and 3-8 for items 2, 3, and 4. Students were therefore more likely to endorse the full range of responses at T2 than at T1, utilizing the full response scale—or nearly so—for all of the EVaCS items at T2.

## Measurement Invariance

Longitudinal measurement invariance of the 14-item three-factor model was tested through a series of nested models. All models were tested in LISREL 8.80 (Jöreskog & Sörbom, 2006) using maximum likelihood estimation with Satorra-Bentler scaled $\chi^2$ and robust standard errors to adjust for multivariate non-normality. The procedure for setting the metric of latent factors is described below.

### Fit Indices

Many of the same indices used to evaluate model fit in Study 1 were also utilized in Study 2. These included the root mean square error of approximation (RMSEA), standardized root mean squared residual (SRMR), and comparative fit index (CFI). Desirable values for these indices when using the Satorra-Bentler adjustment are RMSEA $\leq .05$, SRMR $\leq .07$, and CFI $\geq .96$ (Yu & Muthén, 2002). Local misfit was assessed through correlation residuals of greater than |.10| (Kline, 2011).

The procedure for testing longitudinal invariance examined three nested models—configural, metric, and scalar—where the configural model is nested within the metric and scalar models, and the metric model is nested within the scalar model. Model fit was compared through $\chi^2$ difference tests ($\Delta\chi^2$) as well as how the comparative fit index (CFI) differs between models ($\Delta$CFI). If a $\chi^2$ difference test shows that the more constrained model does not fit significantly worse than the less constrained model (i.e., if $\Delta\chi^2$ is not statistically significant), then the model is said to exhibit at least the level of invariance represented by the more constrained model. For example, if $\chi^2$ for the metric model is not significantly worse (larger) than $\chi^2$ for the configural model, then the model demonstrates metric invariance. A second way to establish invariance is to look at the difference between nested models' CFIs (Cheung & Rensvold, 2002). If the more constrained model's CFI differs from the less constrained model's CFI by .01 or less, then the model is said to exhibit at least the level of invariance represented by the more constrained model. For example, if $CFI_{scalar} = .93$ and $CFI_{metric} = .95$, the measure does not demonstrate scalar invariance because the difference in CFIs is .02. There are differing views as to which index of model fit—either $\Delta\chi^2$ or $\Delta$CFI—is most important for demonstrating invariance (French & Finch, 2006; Little et al., 2007; Steenkamp & Baumgartner, 1998). The $\chi^2$ difference test indicates whether two nested invariance models produce exactly the same fit, and therefore may be overly stringent (similar to the $\chi^2$ test in model fitting) for assessing invariance. However, change in CFI may lack the power to indicate non-invariance in truly non-invariant models (French & Finch, 2006). Therefore, both $\Delta\chi^2$ (as a test of statistical significance) and $\Delta$CFI (as a test of practical significance) were used to evaluate the fit of invariance models, along with correlation residuals.

**Fitting Data from Time Points Separately**

Prior to testing configural invariance, the EVaCS data were fit to T1 and T2 separately. Factor variances were set equal to 1 and factor means equal to zero by default in LISREL. Fit indices are displayed in Table 7. The three-factor model fit well at each time point according to SRMR values of .04 and CFI values of .97; RMSEA (.07) was near the recommended cutoff. The correlation between the Expectancy and Value latent factors was .61 at T1 and .66 at T2. Both factors had negative correlations with Cost (Expectancy: -.55 at T1, -.47 at T2; Value: -.39 at T1, -.40 at T2). All items had factor loadings greater than .70 except item 8, which had a loading of .64 at T1; and item 12, which had a loading of .58 at T1 and .66 at T2. Factor loadings tended to be higher at T2 than at T1. The latent factors accounted for a majority of the variance in most items at both time points ($R^2$ = .51-.77), with the exception of item 8 ($R^2$ = .41-.65) and item 12 ($R^2$ = .34-.43).

Local misfit was assessed via correlation residuals (see Table D4). Of the 91 correlations that were reproduced by the three-factor model, 1 correlation (1%) had a large (> |.10|) residual at T1, and 4 residuals (4%) were large at T2. None of the large residuals exceeded |.15|. Because the model appeared to fit the data from both time points separately, it made sense to test configural invariance by fitting the model to T1 and T2 data simultaneously.

## Scaling the Latent Factors in the Invariance Models

The latent factors were scaled by setting the loading for one item (the referent indicator) for each factor equal to 1 (Little et al., 2007); this was done for each time point (i.e., the configural, metric, and scalar models each had six fixed factor loadings). Items 3 (expectancy), 6 (value), and 15 (cost) were selected as referent indicators. These items were selected because their factor loadings were high at both time points and were comparable in magnitude across time points. It is important to note that this method for scaling the latent factors assumes that items 3, 6, and 15 are invariant across time, and that the value of estimated latent means depends on the particular items that are chosen to serve as referent indicators. In other words, the values of the estimated latent means would change if a different set of referent indicators were used. However, overall model fit would not change if a different set of referent indicators were used, or if a different method were used to scale the latent factors (e.g., fixing variances to 1).

## Configural Invariance

The configural model fit well according to RMSEA (.05), SRMR (.04), and CFI (.98). Factor correlations, loadings, and item $R^2$ values were essentially the same in the configural model as they were when separate models were fit to the time points. The model had six correlation residuals greater than |.10|, representing less than 2 percent of the 378 correlation residuals that were reproduced by the model. Note that in the invariance models, the reproduced covariance matrix includes items from both T1 and T2 and is therefore larger (28×28) than the covariance matrices from Study 1 (15×15 and 14×14). The largest residual was |.16|. The configural model provided good global and local fit to the longitudinal EVaCS data, and therefore served as the baseline model for testing metric invariance.

**Metric Invariance**

Unstandardized pattern coefficients were constrained to be equal across time in the metric model. In addition, intercepts of the six referent indicators (items 3, 6, and 15 at T1 and T2) were set to zero. This model fit well according to global fit indices (RMSEA, .05; SRMR, .04; and CFI, .98). The model had six correlation residuals greater than |.10|, representing less than 2 percent of the 378 correlation residuals that were reproduced by the model. The largest residual was |.15|. The $\chi^2$ difference test between the metric and configural models was statistically significant ($\Delta\chi^2$ (11) = 33.06, $p < .001$), indicating that the metric model fit significantly worse than the configural model. However, because $\Delta$CFI was less than .01 (CFI = .977 in each model), and the metric model continued to produce good global and local fit, it was concluded that the EVaCS sufficiently demonstrated metric invariance. That is, although the metric model fit *statistically* significantly worse than the configural model, for *practical* purposes, the metric model's fit was not substantially worse (see Zilberberg, 2013).

**Scalar Invariance**

To test for scalar invariance, item intercepts (as well as pattern coefficients, constrained in the metric model) were constrained to be equal across time. This model fit well according to SRMR (.04), and CFI (.97); RMSEA (.06) was near but did not reach the cutoff. The scalar model had five correlation residuals greater than |.10|, representing less than 2 percent of the 378 correlation residuals that were reproduced by the model. The largest residual was |.15|. The $\chi^2$ difference test between the metric and configural models was statistically significant ($\Delta\chi^2$ (11) = 310.38, $p < .001$). However, because the change in CFI was less than .01 ($\Delta$CFI = .006), and the metric model continued to produce good global and local fit, it was concluded the EVaCS sufficiently demonstrated scalar invariance. Again, the scalar model fit statistically, but not practically, worse than the metric model.

With scalar invariance established, it made sense to compare observed item means and reproduced (or model-implied) item means. Mean residuals in the scalar model, which are the differences between observed and reproduced item means, were generally small, with an average absolute value of .03. The largest mean residual was |.15| which, on an 8-point response scale, is not substantial.

Parameter and reliability estimates (omega, ω) for the scalar invariant model are shown in Table 8. Across both time points, reliability estimates for the factors were all .79 or higher.

## Latent Mean Differences

Latent mean differences were interpreted through statistical significance tests and standardized latent effect sizes (Table 9). There were no significant differences found for Expectancy ($\Delta\chi^2$ = .08 (1), $p$ = .78) or Value ($\Delta\chi^2$ = .82 (1), $p$ = .37). A statistically significant increase was observed for Cost ($\Delta\chi^2$ = 3.85 (1), $p < .05$), which had a latent Cohen's *d* of .46, representing a medium-sized effect; the latent mean of Cost at T2 is .46 pooled standard deviation units higher than the latent mean of Cost at T1.[12] For the statistically non-significant latent mean differences, Cohen's *d*s were -.06 (for

[12] Latent Cohen's *d* calculated as:

$$\frac{(\kappa_{2j} - \kappa_{1j})}{\sqrt{\psi_{pooled}}}, \text{ where } \sqrt{\psi_{pooled}} = \sqrt{\frac{n_1\psi_{1jj} + n_2\psi_{2jj}}{(n_1 + n_2)}} \times \sqrt{2(1 - r_{12})}$$

Expectancy) and -.19 (for Value). The effect size for Value indicates a small effect; the latent mean of Value at T2 is .19 pooled standard deviation units lower than the latent mean of Value at T1, although this difference is not statistically significant. As shown in Table 9, the statistical significance of and effect sizes for the latent mean differences were similar to those for observed mean differences.[13] This can be attributed to the high internal consistency of the EVaCS scales: Expectancy, $\omega_{T1} = .87$, $\omega_{T2} = .90$; Value, $\omega_{T1} = .89$, $\omega_{T2} = .94$; and Cost, $\omega_{T1} = .79$, $\omega_{T2} = .84$. The scales' high reliability estimates indicate that a large proportion of the variance in observed scores is due to true differences rather than measurement error. Therefore, statistics and effect sizes computed from observed scores will be similar to those computed using latent, "error-free" scores.

**Latent Mean Stability**

The bivariate correlation between a latent factor at two time points—or the *stability coefficient*—represents how much students changed in their rank-ordering on the latent trait from T1 to T2. These latent correlations are displayed in Table 10 along with the observed correlations. The stability coefficient was highest for Cost (.47), followed by Expectancy (.35) and Value (.31). The size of these coefficients indicate that students' rank-ordering on Cost changed moderately—or, remained moderately the same—but rank-ordering fluctuated considerably for Expectancy and Value. Bashkov & Finney (in press) describe how when there is no (or very low) rank-order consistency in scores, there will be no mean-level change because some respondents increased over time

[13] Observed Cohen's *d* calculated as:

$$\frac{(\mu_{2j} - \mu_{1j})}{SD_{pooled}}, \text{ where } SD_{pooled} = \sqrt{\frac{(n_1 - 1)(s_1^2) + (n_2 - 1)(s_2^2)}{(n_1 + n_2 - 2)\sqrt{2(1 - r_{12})}}}$$

while others decreased. Interpreting the latent mean difference and stability results from the present study together, it appears that expectancy and value for coursework did not differ significantly over time at the mean level, likely due to fluctuations in rank-ordering: Some students increased in the traits while other students decreased in the traits, and these changes essentially cancelled each other out so that there were minimal, non-significant mean differences. However, cost for coursework increased significantly over time, and this increase was moderately uniform across students. This means that students overall tended to increase from T1 to T2 in their perceived cost for coursework, and this increase was somewhat consistent for all students, although considerable fluctuation did occur. Variability in the magnitude and direction of individual change over time can be illustrated by plotting individuals' factor scores (on the y-axis) at two or more time points (on the x-axis). Figures 13-15 show the change trajectories for a random sample of 91 students (10% of the matched sample) and show the origins of low stability coefficients and statistically non-significant differences for Expectancy and Value in particular. Table 11 displays individual-level change for the entire matched sample ($N$ = 912). Change was calculated using observed scores by subtracting the total scale score at time point 1 from the total scale score at time point 2. Any negative differences (< 0) are considered decreases, and any positive differences (> 0) are considered increases. A difference of zero from T1 to T2 indicates no change. As shown in Table 11, nearly as many students increased (41%) in Expectancy as decreased (41%). More students decreased (51%) in Value than increased (42%). More students increased (59%) in Cost than decreased (31%).

## Conclusion

By using CFA to test a series of a priori models across independent student samples, the correlated three-factor model was found to demonstrate the best fit to the EVaCS data (research question 1). The model provided good fit in a matched sample across two time points, and configural, metric, and scalar invariance were established for the EVaCS (research question 2). Because the EVaCS demonstrated scalar invariance, latent mean differences were examined for incoming and mid-career students. A statistically and practically significant increase in latent means over time was found for Cost ($d$ = .46), whereas the decreases in Expectancy and Value were not statistically significant (research question 3). Factor stability coefficients indicated considerable variability in students' rank-ordering, pointing to the fact that some students increased while others decreased on the latent factors over time.

# CHAPTER 5.

# DISCUSSION

The purpose of this research was to assess whether the EVaCS is suitable for measuring students' motivation at various time points across the academic career. This was accomplished by examining the scale's psychometric structure and longitudinal invariance in two studies. Study 1 used confirmatory factor analysis to look at the model-data fit of the EVaCS. Study 2 assessed longitudinal invariance and latent mean differences to determine whether the EVaCS functioned similarly in incoming and mid-career student samples, and if so, whether there was a difference in motivation for coursework over time. This research addressed three primary research questions, each of which is discussed below. In the following sections, implications, limitations, and directions for future research are discussed.

## Research Question 1:

## Does the hypothesized three-factor model fit the EVaCS data better than the alternative models tested?

Study 1 tested which of the four a priori models provided the best, parsimonious fit to the EVaCS data. Both the bifactor and correlated three-factor models provided adequate global fit. In the bifactor model, the cost items' loadings were higher on the specific factor (Cost) than on the general factor (Value). Thus, the specific factor accounted for more of the variance in the cost items than did the general factor, which calls into question whether cost is best treated as a value subtype (Wigfield & Eccles, 2000) or if instead it should be treated as a distinct factor (Barron & Hulleman, 2010). Compared to the bifactor model, the three-factor model had fewer sizable correlation residuals and a more favorable AIC value, and it was consistent with the scale developers' intended model. It also had an advantage over the bifactor model of producing a more meaningful, scorable solution—i.e., it is possible to compute distinct means for Expectancy, Value, and Cost. As a result of the analyses, 2 of the original 16 EVaCS items were dropped from the scale (items 11 and 13), resulting in a 14-item scale. The three-factor model with 14 items (4 Expectancy, 6 Value, and 4 Cost) was used for invariance testing in Study 2.

## Research Question 2:

## Does the EVaCS exhibit longitudinal measurement invariance for incoming and mid-career students?

Longitudinal measurement invariance was established in Study 2 by testing a series of configural, metric, and scalar invariance models. Adequate fit was found for the configural model, suggesting that incoming and mid-career students had similar conceptualizations of the EVaCS. The metric model also produced good fit ($\Delta$CFI < .01), indicating that items had equivalent saliency to their respective latent factors at each time point. The fit of the scalar model was not found to be substantially worse than that of the metric model ($\Delta$CFI < .01). Thus, scalar invariance was established, meaning that students with the same level of a latent trait over time will have equal observed means over time. As a result, any change in observed means is therefore due to true change in the level of latent traits.

**Research Question 3:**

**Do the latent means of motivation for coursework differ for incoming and mid-career students?**

A statistically and practically significant difference was found for the latent mean of Cost, which overall increased from time point 1 to time point 2 ($d$ = .46). Furthermore, most students (59%) increased in Cost over time. Slight decreases in latent means were found for Expectancy ($d$ = -.06) and Value ($d$ = -.19) overall; however, neither were statistically significant. Individual-level change analyses (Table 11 and Figures 13 and 14) indicated that many students increased in expectancy and value while many others decreased.

## Implications for Measuring Motivation with the EVaCS

By establishing the EVaCS' psychometric structure and longitudinal measurement invariance in multiple undergraduate student samples, this research has provided strong support for the EVaCS' use in measuring students' motivation for coursework in a semester. Recall that the EVaCS measures expectancy, value, and cost at a *general* level for semester coursework. Because it measures these general constructs, the EVaCS has the advantage of being a relatively short, easily scored instrument, which may promote its adoption by educators and researchers who are only interested in measuring these constructs at a general level. However, for studies that emphasize a particular subtype of a construct—e.g., utility value—a general measure like the EVaCS may be inadequate. In the present research, many of the hypotheses about how motivation may change across the academic career referred to students' perceived utility value for coursework. It is possible that latent mean differences would have been observed if a measure of utility value had been administered as opposed to the EVaCS' general value subscale. It is important for researchers and practitioners to evaluate whether a general measure of expectancy, value, and cost will be sufficient for their purposes, or whether a more nuanced measure of these constructs is needed.

Although the scale is currently usable for many purposes, additional work can be done to improve the EVaCS psychometrically. For example, items 11 and 13, deemed problematic in Study 1, should be removed, revised, and/or replaced to ensure that the breadth of the cost construct is covered. Item 13 ("I think there will be other things I'd rather do with my time than just focusing on my classes") was excluded from analyses at the outset due to its negative effect on scale reliability. It is questionable whether this item is indicative of cost at all: Simply wanting to do other things besides focusing on classes does not imply cost if those other activities will not interfere with one's ability to do schoolwork. Item 11 was removed from the models after it was found to be associated with large correlation residuals and only moderate factor loadings in Samples 1 and 2. This item ("I think my classes will require too much time for me to do well") is an example of time/effort-related cost. With its removal from the EVaCS, only one time/effort-related cost item remained (item 14), along with three items about other activities that may interfere with being successful in classes (items 12, 15, and 16). Some minor local misfit between item 14 and the expectancy items at time point 2 persisted through the invariance models, possibly due to the fact that item 14's content was different than the other cost items' content. Therefore, even if the breadth of the cost construct is adequately covered by items 12, 14, 15, and 16, their differential content may still contribute to poor fit for the Cost factor. The EVaCS does not appear to cover the full breadth of cost, though, because none of the original 16 items made reference to psychological aspects (e.g., stress, anxiety) that are theorized to impact perceived cost. More cost items may need to be piloted to determine which are best able to represent the breadth of cost in a general sense for the EVaCS.

The positioning of cost items on the EVaCS should be carefully considered as well. It is interesting that the three most problematic items in Study 1 in terms of corresponding to the hypothesized factor structure were the first three cost items (11-13). Because the cost items are framed in such a way that a highly motivated student would likely select a *low* response—unlike high responses to the expectancy and value items—it is possible that their appearance all together at the end of the scale requires additional cognitive effort to process (see Appendix A). If students do not immediately shift their orientation to the items, their responses to the early cost items may be more prone to error than later item responses. To explore this idea, the EVaCS could be administered to a sample of students with the cost items either reordered at the end of the scale or appearing in amongst the expectancy and value items. It is possible that items 11, 12, and 13 may fit the three-factor model better when repositioned on the scale.

## Implications for Motivation Theory

Although it was not intended as the primary purpose of this research, a major aspect of Study 1 was investigating the appropriate placement of cost in the expectancy-value framework. Results from both the bifactor and three-factor models supported the idea that cost items were distinct enough from the EVaCS value items to warrant positioning Cost as a separate construct in a CFA model. By separating cost from value, it was possible to obtain a factor correlation between Expectancy and Cost. The relationship between expectancy and cost is interesting, particularly in how it differs from the relationship between expectancy and value. Attainment, intrinsic, and utility value—which theoretically comprise the general value factor—seem to bear a moderately strong positive relationship with expectancy either because people value what they expect to be successful at, or they expect to be successful at the activities they value. The story is different with expectancy and cost, which have a moderately negative relationship. Students with high expectancies may require less effort to be successful at a task, so the effort expended is not high enough to be perceived as a cost. Conversely, students with low expectancies may perceive greater cost (i.e., require more effort) in order to be successful. Therefore, when cost items are oriented toward expending effort, it makes sense that there will be a negative relationship between expectancy and cost. Only one item in the final 14-item EVaCS asked specifically about effort—item 14, "I don't think I can invest the time and effort that is needed to do well in my classes." This was the item primarily associated with local misfit in many of the invariance models due to its large correlation residuals with expectancy items at the second time point. The reason for this may have been that, as the lone effort-related item on the cost scale, item 14 was less "like" the other non-effort-related cost items and therefore was less represented by Cost factor (as indicated by the lower item $R^2$ values). This point is important to consider because the relative (im)balance of effort- and non-effort-related cost items can have implications for how cost empirically relates to expectancy and value—e.g., whether it is best treated as a value subtype or independent factor. In addition to the theoretical implications of using effort-related cost items, it is important to note that items with affective content (e.g., stress, fear of failure) were not included on the EVaCS, and their representation on a scale may also influence empirical relationships among constructs.

**Limitations of the Studies**

The premise of this research was that students' motivation—in particular, their value—may change over time as students take different courses each semester. Students take primarily general education courses early in their academic careers and primarily major courses later on. However, the two time points investigated in Study 2 were (1) college entry and (2) midway through the second semester of the second year of college. One limitation of this study is that it is unclear to what extent the composition of coursework (in terms of general education vs. in-major) differs for these two time points. It is likely that, because both measurement occasions occurred in the first two years of college, students were taking several general education courses at *both* of these time points. Therefore, the composition of coursework may not have been different enough to produce mean-level changes in value. In addition, students may have changed in other (unmeasured) variables at these two time points, such as their major or career decidedness, which influenced their motivation for coursework. Additional variables should be studied to provide more context for why change in motivation for coursework does or does not occur (e.g., attitudes toward general education, reasons for attending college). Such studies could contribute external validity evidence for the EVaCS if the variables under investigation relate to EVaCS scores in theoretically predicted ways.

Only two time points were looked at in this study, and it is unclear whether mean-level motivation for coursework may change in a linear as opposed to a nonlinear fashion if measured at additional time points. That is, students on average may not have changed significantly in value from college entry to midway through their fourth semester, but value for coursework may increase later in the academic career. In other words, an absence of change in motivation across the two time points examined may represent temporary rather than continual stability in mean-level motivation. Because at least three time points are needed to model growth trends over time, more time points should be investigated to better understand if and how motivation for coursework changes across the academic career. In addition, the invariance of the EVaCS should be examined across later time points.

Another limitation concerns the timing of the first data collection time point, which occurred a few days prior to students' first week of college courses. It is possible that, having not yet experienced college coursework firsthand, students were somewhat naïve in their self-reports of expectancy, value, and cost in the higher education context. Thus, it is not clear to what extent the levels of motivation reported by students prior to the start of classes may have been over- or understated due to lack of experience in the postsecondary academic environment. It is important to recognize that comparisons between motivation at college entry and midway through college may be impacted by postsecondary inexperience in a way that other cross-time comparisons (e.g., junior to senior year) are not. The potential impact of postsecondary inexperience on incoming students' reported motivation for coursework could be examined by administering the EVaCS to a sample of students immediately prior to their first semester of college and again a few weeks into their first semester. If there are substantial differences in incoming students' motivation over a short duration of time, and if similar differences are not observed over short durations later in the academic career, this would suggest that incoming students' responses are somewhat influenced by postsecondary inexperience.

Finally, unlike the CFAs in Study 1 that were cross-validated on independent samples, the fit of the invariance models in Study 2 were not cross-validated. Therefore, the EVaCS' measurement invariance needs to be replicated in another longitudinal sample. The generalizability of the findings presented here would be strengthened if they were replicated at other higher education institutions with different characteristics and student demographics.

## Future Research on Motivation for Coursework

As was emphasized in Chapter 1, the rationale for undertaking this research stemmed from the basic idea that motivation for coursework depends on the particular courses students are taking in a semester, and because courses change each semester, so too should students' motivation. Hypotheses were offered as to why or why not changes in value might be observed due to such things as the shift in emphasis from general education to major coursework; students' major and career decidedness; their interest in acquiring a broad, general education, and their attitudes toward general education. The current research found no statistically significant mean difference over time in students' value for coursework. However, stability coefficients and an examination of change at the individual-level revealed that change in value at the mean level was impacted by individual differences: Some students' increases were counteracted by other students' decreases, which contributed to no mean difference in value. The same was true for students' expectancy for their coursework, which did not differ significantly over time at the mean-level and, like value, had many students' increases matched by other students' decreases.

Thus, the question remains as to why some students were increasing whereas others were decreasing in expectancy and value over time. The reasons for these different patterns of change were not explored in the present research, but there are a number of possible explanations that could be explored in the future. For example, increases in expectancy may have been observed for students who performed well in courses their first few semesters and therefore developed higher expectations for success than they had as incoming freshmen. Decreases in expectancy may have been observed for students whose courseloads were perceived as less difficult at college entry than midway through the second year of college. Increases in value might have occurred for students who had begun taking courses in their major midway through the second year of college but were not taking major coursework in their first semester of college. Decreases in value might have occurred among students who were taking a high proportion of general education courses at both time points, but who had developed more negative attitudes toward general education during their time at the university, perhaps owing to ineffective instruction. That is, some incoming freshmen may have favorable attitudes toward general education *until* they take these courses or interact with others who have negative attitudes.

This research found a statistically significant latent mean increase in cost; however, understanding the reasons behind the mean-level change in cost requires additional research. For instance, the general trend toward higher perceived cost may be due to students' increased involvement in extracurricular activities (e.g., employment, student organizations, socializing) as they proceed through college. In other words, mid-career students may experience more competing demands for their time than incoming freshmen do. Another possible explanation is that for some students, the increase in perceived cost involved wanting to focus on coursework in the major rather than fulfilling general education requirements. Several of the scholarly writings and student anecdotes presented in Chapters 1 and 2 highlighted the opinion that general education courses take time away from major coursework. Thus, there is cost associated with general education coursework because students perceive that they are "missing out" on other, more appealing academic opportunities. If this were the case, and if students' vocational focus increases over time, it may follow that their perceived cost increases across semesters in which they are taking several general education courses.

These are merely hypothetical explanations intended to highlight the different patterns of change in expectancy, value, and cost for individual students. Their legitimacy would need to be substantiated through further research. Based on findings from the present research, the EVaCS can be used with confidence to pursue questions of why motivation for coursework does or does not change over time. Because the EVaCS demonstrated a theoretically-supported structure and longitudinal invariance across two time points, scale scores can be the dependent variables in studies aimed at understanding not only what variables contributes to motivation at a *single* time point but also what contributes to *change* in motivation over time. For instance, scores from the cost scale could be correlated with the proportion of general education credits that students take in a semester. Another question to explore is whether students who enter college with undecided majors differ more in their value for coursework over time than students with decided majors. As previously discussed, undecided students may highly value early general education coursework which enables them to explore the curriculum, *and* value the major coursework they take later in their academic careers. Consequently, their motivation for coursework in a semester may not change as much as decided students who may value general education coursework far less than their major coursework.

Although the EVaCS has demonstrated several positive qualities, in order to best address some of the ideas introduced by this research—e.g., students' perceived utility value for general education in terms of their future careers—a different or modified measure is needed. Greater emphasis would need to be placed on utility value because currently the EVaCS contains only one utility value item (item 7, "I think my classes will be useful"). A deeper understanding of how utility value contributes to students' motivation for general education could be key for designing and implementing interventions aimed at increasing their utility value for this coursework—which, for some students, is minimal. In addition, the EVaCS items and instructions would need to be modified so that they are more specific to general education. The focus of a modified scale could be the general education curriculum broadly, or specific general education course(s). For example, the instructions might begin with, "For this survey we are interested in your attitudes regarding *general education coursework*," with an item like "I think my general education classes will be useful." Grays et al.'s (2012) research on students' motivation for general education, while relevant and informative, was limited by the fact that their motivation instrument lacked construct breadth. Although there are advantages to using short instruments, particularly in terms of administration, such research can be enhanced by using an instrument such as the EVaCS (with modifications) to more comprehensively assess motivation for different components of the general education curriculum. Other measures (e.g., attitudes toward general education, reasons for attending college, and career decidedness) would also be needed in order to fully examine students' motivation for coursework (i.e., Benson's (1998) external stage).

## Assessing Students' Motivation in Higher Education

Writing specifically about general education, Johnston et al. (1991) noted how "little has been done to identify and understand student perceptions regarding general education, much less reshape and harness them on its behalf" (p. 182). One step toward understanding and reshaping student motivation—not only for general education, but all postsecondary coursework—is the establishment of an instrument that can yield valid inferences about student motivation. Because of its theoretically- and psychometrically-supported structure, the EVaCS can be administered as a short, reliable measure of motivation for semester coursework by higher education institutions wishing to better understand students' expectancy, value, and cost. Data obtained through the EVaCS may prompt the creation of more refined measures to fit specific purposes (e.g., measuring motivation for a particular course or area of the general education curriculum) or more focused research on the factors that contribute to students' motivation for their courses. If adapted for use in a specific course, the EVaCS may be useful in identifying students' motivational barriers to achieving the intended learning outcomes. For instance, students may not expect to be successful in a course that is known for tough grading and time-intensive assignments, or they may be bored by a course whose content seems too far removed from their everyday lives. The EVaCS may also identify educational policies and practices that are effective in motivating students, such as activities to enhance utility value in general education courses.

While research on student motivation and the factors that influence it can make meaningful contributions to educational research, theory, and practice, it is vital for such research to ultimately make connections to student learning outcomes. With an intense emphasis on effectiveness and accountability at all levels of education, academic institutions cannot ignore the powerful role that motivation plays in students' success. There are abundant theoretical arguments and empirical evidence which make clear that motivation positively impacts student learning. If institutions are to work to intentionally impact student motivation in order to produce improved learning outcomes, they must understand their students' motivation. In order for institutions to understand their students' motivation, they must measure it. The present research supports the use of the EVaCS as a general measure of motivation for college coursework. It is hoped that the EVaCS will be adopted and administered by institutions to provide a more comprehensive picture of how motivation impacts learning outcomes, and how motivation can be "harnessed" to promote students' achievement of learning outcomes.

## Conclusion

This research established a theoretically and psychometrically-supported scale for measuring and assessing mean differences over time in college students' motivation for semester coursework—the Expectancy, Value, and Cost Scale (EVaCS). Additional studies may be conducted to further improve the scale and to better understand what contributes to motivation for coursework at a single time point or across multiple time points. Ultimately, it is hoped that the EVaCS will prove useful in identifying and addressing students' motivational challenges so that intended learning outcomes will be successfully achieved.

## Appendix A

Expectancy, Value, and Cost Scale (EVaCS)

For this survey we are interested in your *general*, *overall* attitudes regarding all of the classes you have this semester. Please read each item and choose the response choice, using the 1 to 8 scale below, that best represents your feelings about how true each item is. If you *Completely Disagree* with the statement, mark a 1. If you *Completely Agree* with the statement, mark an 8. Or mark any number in between. There are no right or wrong answers. Just answer as honestly as possible.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| Completely disagree | Strongly disagree | Disagree | Slightly disagree | Slightly agree | Agree | Strongly agree | Completely agree |

**Expectancy items**

1. I expect to do well in my classes.
2. I am confident that I can learn the material in my classes.
3. I am confident I will be successful in my classes.
4. I know I can understand the material in my classes.

**Value items**

5. I value the classes I am taking this semester.
6. I think my classes will be worthwhile.
7. I think my classes will be useful.
8. I think my classes will be interesting.
9. I see a purpose for taking my classes this semester.
10. I see why my classes are important to take.

**Cost items**

11. I think my classes will require too much time for me to do well.
12. Because of the other things that I want to do in college, I don't think I'll have as much time to put into my classes as I'd like.
13. I think there will be other things I'd rather do with my time than just focusing on my classes.
14. I don't think I can invest the time and effort that is needed to do well in my classes.
15. Doing well in my classes may not be worth all the things I have to give up.
16. Because of other things I'm interested in, I'm not sure I want to sacrifice what will be needed to do well in my classes.

**Order of administration:** 5, 1, 6, 2, 7, 3, 8, 9, 4, 10, 11, 12, 13, 14, 15, 16

# Appendix B

Study 1 Item Correlations and Descriptive Statistics

Table B1

*Item Correlations and Descriptive Statistics for Sample 1, N = 921*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1. | | | | | | | | | | | | | | |
| 2 | .59 | 1. | | | | | | | | | | | | | |
| 3 | .69 | .68 | 1. | | | | | | | | | | | | |
| 4 | .52 | .73 | .65 | 1. | | | | | | | | | | | |
| 5 | .34 | .37 | .35 | .36 | 1. | | | | | | | | | | |
| 6 | .43 | .46 | .44 | .41 | .70 | 1. | | | | | | | | | |
| 7 | .34 | .43 | .40 | .43 | .64 | .75 | 1. | | | | | | | | |
| 8 | .34 | .41 | .44 | .46 | .52 | .61 | .62 | 1. | | | | | | | |
| 9 | .34 | .41 | .42 | .45 | .65 | .68 | .71 | .59 | 1. | | | | | | |
| 10 | .31 | .38 | .39 | .43 | .59 | .65 | .68 | .57 | .74 | 1. | | | | | |
| 11 | -.27 | -.33 | -.31 | -.29 | -.12 | -.20 | -.18 | -.19 | -.22 | -.16 | 1. | | | | |
| 12 | -.28 | -.26 | -.29 | -.24 | -.18 | -.21 | -.22 | -.19 | -.22 | -.19 | .52 | 1. | | | |
| 14 | -.39 | -.39 | -.40 | -.34 | -.29 | -.31 | -.29 | -.26 | -.33 | -.31 | .49 | .61 | 1. | | |
| 15 | -.29 | -.25 | -.29 | -.26 | -.27 | -.28 | -.28 | -.24 | -.32 | -.30 | .34 | .49 | .63 | 1. | |
| 16 | -.30 | -.23 | -.26 | -.25 | -.28 | -.25 | -.25 | -.22 | -.28 | -.28 | .31 | .50 | .60 | .70 | 1. |
| *SD* | 0.92 | 0.94 | 0.96 | 1.03 | 1.06 | 1.02 | 1.10 | 1.14 | 1.13 | 1.14 | 1.54 | 1.60 | 1.29 | 1.36 | 1.46 |
| Mean | 6.76 | 6.66 | 6.54 | 6.45 | 6.74 | 6.56 | 6.55 | 6.33 | 6.56 | 6.49 | 3.61 | 3.71 | 2.71 | 2.61 | 2.86 |
| Skew | -0.12 | -0.25 | -0.15 | -0.09 | -0.72 | -0.43 | -0.52 | -0.52 | -0.76 | -0.67 | 0.73 | 0.47 | 0.74 | 0.84 | 0.84 |
| Kurt | -0.85 | -0.21 | -0.38 | -0.57 | 1.25 | 0.43 | 0.34 | 0.64 | 1.11 | 0.87 | 0.49 | -0.14 | 0.89 | 0.62 | 0.81 |

Table B2

*Item Correlations and Descriptive Statistics for Sample 2, N = 923*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1. | | | | | | | | | | | | | | |
| 2 | .64 | 1. | | | | | | | | | | | | | |
| 3 | .70 | .72 | 1. | | | | | | | | | | | | |
| 4 | .51 | .70 | .65 | 1. | | | | | | | | | | | |
| 5 | .39 | .43 | .39 | .39 | 1. | | | | | | | | | | |
| 6 | .47 | .49 | .50 | .44 | .70 | 1. | | | | | | | | | |
| 7 | .40 | .46 | .46 | .40 | .61 | .76 | 1. | | | | | | | | |
| 8 | .34 | .42 | .46 | .45 | .56 | .61 | .60 | 1. | | | | | | | |
| 9 | .38 | .47 | .46 | .46 | .64 | .68 | .71 | .61 | 1. | | | | | | |
| 10 | .35 | .44 | .43 | .46 | .56 | .65 | .67 | .53 | .75 | 1. | | | | | |
| 11 | -.25 | -.31 | -.33 | -.32 | -.15 | -.19 | -.19 | -.23 | -.22 | -.19 | 1. | | | | |
| 12 | -.24 | -.26 | -.29 | -.22 | -.18 | -.24 | -.25 | -.22 | -.23 | -.23 | .51 | 1. | | | |
| 14 | -.39 | -.42 | -.43 | -.41 | -.27 | -.31 | -.30 | -.27 | -.33 | -.31 | .52 | .54 | 1. | | |
| 15 | -.32 | -.32 | -.33 | -.27 | -.30 | -.31 | -.32 | -.29 | -.34 | -.32 | .38 | .52 | .63 | 1. | |
| 16 | -.22 | -.28 | -.27 | -.22 | -.24 | -.25 | -.26 | -.25 | -.27 | -.27 | .36 | .46 | .55 | .68 | 1. |
| | | | | | | | | | | | | | | | |
| *SD* | 0.94 | 0.96 | 1.01 | 1.06 | 1.02 | 1.04 | 1.11 | 1.14 | 1.10 | 1.15 | 1.42 | 1.61 | 1.26 | 1.32 | 1.46 |
| Mean | 6.77 | 6.71 | 6.55 | 6.46 | 6.75 | 6.58 | 6.57 | 6.35 | 6.61 | 6.56 | 3.58 | 3.66 | 2.72 | 2.56 | 2.84 |
| Skew | -0.22 | -0.35 | -0.15 | -0.29 | -0.62 | -0.44 | -0.42 | -0.25 | -0.55 | -0.61 | 0.68 | 0.50 | 0.79 | 0.89 | 0.81 |
| Kurt | -0.70 | -0.23 | -0.58 | -0.14 | 0.74 | 0.14 | -0.17 | -0.40 | 0.28 | 0.47 | 0.73 | -0.03 | 0.98 | 1.04 | 0.61 |

Table B3

*Item Correlations and Descriptive Statistics for Sample 3, N = 916*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1. | | | | | | | | | | | | | | |
| 2 | .63 | 1. | | | | | | | | | | | | | |
| 3 | .72 | .70 | 1. | | | | | | | | | | | | |
| 4 | .58 | .74 | .69 | 1. | | | | | | | | | | | |
| 5 | .41 | .43 | .40 | .45 | 1. | | | | | | | | | | |
| 6 | .45 | .46 | .46 | .43 | .72 | 1. | | | | | | | | | |
| 7 | .41 | .45 | .45 | .42 | .67 | .76 | 1. | | | | | | | | |
| 8 | .41 | .44 | .48 | .47 | .61 | .68 | .63 | 1. | | | | | | | |
| 9 | .39 | .45 | .45 | .48 | .65 | .71 | .71 | .66 | 1. | | | | | | |
| 10 | .37 | .42 | .44 | .47 | .61 | .65 | .69 | .61 | .77 | 1. | | | | | |
| 11 | -.34 | -.38 | -.42 | -.36 | -.20 | -.24 | -.21 | -.27 | -.24 | -.24 | 1. | | | | |
| 12 | -.26 | -.24 | -.28 | -.23 | -.19 | -.26 | -.21 | -.24 | -.23 | -.22 | .53 | 1. | | | |
| 14 | -.35 | -.39 | -.44 | -.37 | -.27 | -.31 | -.29 | -.30 | -.32 | -.32 | .60 | .62 | 1. | | |
| 15 | -.31 | -.30 | -.33 | -.26 | -.28 | -.33 | -.29 | -.28 | -.32 | -.29 | .41 | .50 | .61 | 1. | |
| 16 | -.29 | -.28 | -.32 | -.26 | -.22 | -.26 | -.25 | -.23 | -.24 | -.26 | .44 | .55 | .62 | .67 | 1. |
| | | | | | | | | | | | | | | | |
| *SD* | 0.94 | 0.98 | 1.03 | 1.08 | 1.05 | 1.08 | 1.11 | 1.19 | 1.20 | 1.18 | 1.40 | 1.67 | 1.35 | 1.43 | 1.49 |
| Mean | 6.73 | 6.71 | 6.56 | 6.41 | 6.73 | 6.54 | 6.54 | 6.32 | 6.57 | 6.48 | 3.59 | 3.75 | 2.75 | 2.60 | 2.86 |
| Skew | -0.38 | -0.65 | -0.40 | -0.45 | -0.61 | -0.53 | -0.56 | -0.44 | -0.76 | -0.61 | 0.62 | 0.40 | 0.89 | 1.05 | 0.75 |
| Kurt | 0.20 | 1.15 | 0.16 | 0.38 | 0.57 | 0.38 | 0.26 | 0.07 | 0.54 | 0.36 | 0.46 | -0.36 | 1.23 | 1.36 | 0.29 |

Table B4

*Item Correlations and Descriptive Statistics for Sample 4, N = 922*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1. | | | | | | | | | | | | | | |
| 2 | .60 | 1. | | | | | | | | | | | | | |
| 3 | .68 | .71 | 1. | | | | | | | | | | | | |
| 4 | .50 | .66 | .65 | 1. | | | | | | | | | | | |
| 5 | .40 | .43 | .44 | .38 | 1. | | | | | | | | | | |
| 6 | .45 | .47 | .48 | .42 | .72 | 1. | | | | | | | | | |
| 7 | .38 | .49 | .45 | .38 | .66 | .73 | 1. | | | | | | | | |
| 8 | .40 | .45 | .51 | .46 | .56 | .62 | .59 | 1. | | | | | | | |
| 9 | .38 | .47 | .45 | .42 | .66 | .69 | .72 | .62 | 1. | | | | | | |
| 10 | .37 | .44 | .44 | .43 | .64 | .66 | .70 | .58 | .79 | 1. | | | | | |
| 11 | -.25 | -.33 | -.34 | -.34 | -.18 | -.19 | -.17 | -.19 | -.18 | -.19 | 1. | | | | |
| 12 | -.26 | -.24 | -.30 | -.25 | -.20 | -.20 | -.19 | -.22 | -.18 | -.19 | .53 | 1. | | | |
| 14 | -.42 | -.39 | -.45 | -.32 | -.30 | -.30 | -.30 | -.31 | -.30 | -.30 | .47 | .54 | 1. | | |
| 15 | -.29 | -.25 | -.28 | -.20 | -.32 | -.31 | -.28 | -.26 | -.28 | -.31 | .30 | .42 | .49 | 1. | |
| 16 | -.31 | -.25 | -.29 | -.22 | -.30 | -.29 | -.26 | -.26 | -.26 | -.31 | .31 | .47 | .53 | .63 | 1. |
| | | | | | | | | | | | | | | | |
| *SD* | 0.92 | 0.96 | 1.00 | 1.06 | 1.04 | 1.04 | 1.10 | 1.12 | 1.15 | 1.20 | 1.46 | 1.62 | 1.26 | 1.36 | 1.40 |
| Mean | 6.73 | 6.65 | 6.49 | 6.42 | 6.73 | 6.59 | 6.57 | 6.32 | 6.58 | 6.51 | 3.66 | 3.81 | 2.78 | 2.64 | 2.86 |
| Skew | -0.17 | -0.24 | -0.19 | -0.24 | -0.52 | -0.45 | -0.60 | -0.37 | -0.72 | -0.62 | 0.80 | 0.46 | 0.69 | 0.99 | 0.65 |
| Kurt | -0.62 | -0.49 | -0.39 | -0.37 | 0.13 | 0.11 | 0.40 | -0.10 | 0.99 | 0.32 | 0.76 | -0.16 | 0.72 | 1.46 | 0.19 |

# Appendix C

## Study 1 Correlation Residual Matrices

Table C1

*Correlation Residuals for the 15-Item Bifactor Model*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | -.01 | .05 | -.07 | .01 | .04 | -.03 | -.02 | -.05 | -.05 | **-.13** | -.10 | **-.19** | **-.12** | -.06 |
| 2 | -.02 | | -.02 | .04 | .00 | .00 | -.02 | .01 | -.01 | -.01 | **-.17** | -.10 | **-.20** | -.09 | -.10 |
| 3 | .08 | -.02 | | -.01 | -.05 | .01 | -.03 | .05 | -.02 | -.03 | **-.18** | **-.12** | **-.20** | **-.11** | -.08 |
| 4 | -.06 | .05 | -.02 | | .00 | .01 | -.03 | .09 | .03 | .05 | **-.19** | -.07 | **-.21** | -.07 | -.05 |
| 5 | -.01 | -.03 | -.05 | -.03 | | .05 | -.03 | .03 | .00 | -.04 | .05 | .04 | .03 | .00 | .00 |
| 6 | .05 | .01 | .00 | -.01 | .05 | | .04 | .00 | -.05 | -.03 | .03 | .01 | .02 | .03 | .03 |
| 7 | -.04 | -.01 | -.04 | .00 | -.01 | .03 | | .00 | -.01 | -.01 | .02 | .00 | .03 | .01 | .01 |
| 8 | .01 | .04 | .08 | .10 | -.03 | .00 | .01 | | .01 | -.04 | -.05 | -.01 | .01 | .00 | -.02 |
| 9 | -.03 | -.03 | -.02 | .03 | .01 | -.04 | -.01 | -.01 | | .08 | .00 | .01 | .00 | .00 | .00 |
| 10 | -.05 | -.04 | -.02 | .02 | -.02 | -.03 | -.01 | .00 | .07 | | .02 | .00 | .00 | .00 | -.01 |
| 11 | **-.16** | **-.21** | .01 | **-.17** | .06 | .00 | .02 | -.02 | -.02 | .03 | | **.13** | .08 | -.08 | -.07 |
| 12 | **-.16** | **-.13** | **-.16** | **-.11** | .02 | .01 | .01 | .00 | .00 | .02 | **.14** | | .02 | -.03 | -.04 |
| 14 | **-.22** | **-.19** | **-.21** | **-.15** | .00 | .02 | .04 | .01 | -.01 | .00 | .05 | .03 | | -.01 | -.03 |
| 15 | **-.13** | -.07 | -.10 | -.08 | .01 | .02 | .02 | .01 | -.02 | -.01 | -.08 | -.06 | -.02 | | .07 |
| 16 | **-.15** | -.06 | -.09 | -.09 | -.03 | .03 | .03 | .01 | .00 | -.01 | -.10 | -.04 | -.03 | .10 | |

*Note.* Sample 1 below the main diagonal, Sample 2 above the main diagonal. Residuals larger than |.10| are in boldface.

Table C2

*Correlation Residuals for the 15-Item Three-Factor Model*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | -.01 | .05 | -.07 | .01 | .04 | -.02 | -.01 | -.04 | -.04 | -.03 | .02 | -.07 | .00 | .07 |
| 2 | -.02 | | -.02 | .04 | .01 | .01 | -.02 | .02 | -.01 | -.01 | -.06 | .04 | -.06 | .04 | .05 |
| 3 | .07 | -.02 | | -.01 | -.04 | .01 | -.02 | .06 | -.02 | -.02 | -.07 | .01 | -.06 | .03 | .07 |
| 4 | -.07 | .05 | -.02 | | .01 | .01 | -.02 | .09 | .03 | .06 | -.09 | .05 | -.09 | .05 | .08 |
| 5 | -.01 | -.02 | -.04 | -.02 | | .05 | -.03 | .03 | .00 | -.04 | .05 | .04 | .01 | -.02 | .01 |
| 6 | .05 | .02 | .01 | .00 | .05 | | .04 | .00 | -.05 | -.03 | .03 | .01 | .00 | .00 | .04 |
| 7 | -.04 | .00 | -.03 | .01 | -.01 | .03 | | -.01 | -.01 | -.01 | .03 | .01 | .01 | -.01 | .02 |
| 8 | .02 | .05 | .08 | .10 | -.03 | .00 | .01 | | .01 | -.04 | -.05 | -.01 | -.01 | -.03 | -.01 |
| 9 | -.03 | -.02 | -.01 | .04 | .01 | -.04 | -.01 | -.01 | | .08 | .00 | .02 | -.02 | -.03 | .01 |
| 10 | -.05 | -.04 | -.02 | .03 | -.02 | -.03 | -.01 | .00 | .07 | | .02 | .01 | -.02 | -.03 | -.01 |
| 11 | -.07 | -.10 | -.08 | -.07 | .06 | -.01 | .02 | -.03 | -.02 | .02 | | **.13** | .06 | -.09 | -.07 |
| 12 | -.02 | .03 | .00 | .04 | .05 | .04 | .04 | .02 | .03 | .05 | **.13** | | .00 | -.02 | -.03 |
| 14 | -.09 | -.04 | -.05 | -.01 | -.02 | .00 | .02 | .00 | -.03 | -.02 | .03 | .02 | | -.02 | -.04 |
| 15 | -.01 | .07 | .03 | .05 | -.02 | .00 | .00 | -.01 | -.05 | -.04 | -.08 | -.05 | -.02 | | .09 |
| 16 | -.02 | .08 | .05 | .05 | -.04 | .02 | .02 | .01 | -.01 | -.02 | -.10 | -.03 | -.03 | **.12** | |

*Note.* Sample 1 below the main diagonal, Sample 2 above the main diagonal. Residuals larger than |.10| are in boldface.

Table C3

*Correlation Residuals for the 14-Item Bifactor Model*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | -.02 | .04 | -.05 | .01 | .04 | -.03 | .05 | -.04 | -.04 | **-.14** | **-.23** | **-.11** | **-.14** |
| 2 | -.02 | | -.01 | .04 | .00 | .01 | .03 | .06 | -.01 | -.02 | **-.11** | **-.18** | -.05 | -.06 |
| 3 | .06 | -.02 | | .00 | -.02 | .00 | -.03 | .10 | -.04 | -.04 | **-.16** | **-.23** | -.07 | -.09 |
| 4 | -.05 | .04 | -.02 | | -.02 | .00 | -.04 | .10 | -.01 | .01 | **-.13** | **-.13** | -.02 | -.04 |
| 5 | .01 | .00 | -.03 | .03 | | .05 | .00 | -.01 | -.02 | -.02 | .00 | .00 | -.03 | -.02 |
| 6 | .02 | -.01 | -.02 | -.03 | .04 | | .03 | .02 | -.03 | -.04 | .01 | .02 | .00 | .00 |
| 7 | -.01 | -.02 | -.02 | -.03 | .00 | .03 | | -.02 | .00 | .00 | .02 | .02 | .03 | .03 |
| 8 | .03 | .02 | .05 | .06 | .01 | .01 | -.02 | | .00 | -.02 | -.04 | -.03 | .00 | .00 |
| 9 | -.04 | -.02 | -.02 | .03 | -.02 | -.02 | -.01 | .00 | | .07 | .03 | .03 | .03 | .04 |
| 10 | -.04 | -.03 | -.01 | .04 | -.03 | -.05 | .00 | -.01 | .08 | | .02 | .02 | -.01 | -.02 |
| 12 | **-.12** | -.08 | **-.13** | -.08 | .03 | -.02 | .03 | -.02 | .01 | .01 | | **.11** | -.05 | -.03 |
| 14 | **-.16** | **-.19** | **-.22** | **-.17** | .03 | .02 | .03 | -.01 | .00 | -.01 | .06 | | -.03 | -.03 |
| 15 | **-.12** | -.10 | **-.13** | -.06 | .02 | -.01 | .03 | .01 | -.01 | .01 | -.05 | -.01 | | .04 |
| 16 | **-.14** | **-.11** | **-.15** | -.10 | .02 | .01 | .01 | .01 | .02 | -.01 | -.02 | -.02 | .04 | |

*Note.* Sample 3 below the main diagonal, Sample 4 above the main diagonal. Residuals larger than |.10| are in boldface.

Table C4

*Correlation Residuals for the 14-Item Three-Factor Model*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | -.02 | .03 | -.05 | .01 | .04 | -.03 | .05 | -.04 | -.04 | -.02 | **-.14** | -.01 | -.01 |
| 2 | -.02 | | -.01 | .04 | .01 | .01 | .04 | .06 | .00 | -.01 | .02 | -.08 | .05 | .07 |
| 3 | .05 | -.02 | | .00 | -.01 | .01 | -.03 | .10 | -.04 | -.04 | -.02 | **-.12** | .05 | .05 |
| 4 | -.05 | .05 | -.02 | | -.01 | .01 | -.03 | **.11** | .00 | .02 | -.01 | -.04 | .08 | .08 |
| 5 | .02 | .01 | -.03 | .04 | | .05 | .00 | -.01 | -.02 | -.02 | .03 | -.03 | -.06 | -.02 |
| 6 | .03 | .00 | -.02 | -.02 | .04 | | .03 | .02 | -.03 | -.04 | .04 | -.02 | -.03 | .01 |
| 7 | -.01 | -.01 | -.02 | -.03 | .00 | .03 | | -.02 | -.01 | -.01 | .05 | -.02 | .00 | .03 |
| 8 | .03 | .03 | .05 | .07 | .01 | .01 | -.02 | | .00 | -.02 | -.01 | -.06 | -.02 | .00 |
| 9 | -.03 | -.01 | -.02 | .03 | -.02 | -.02 | -.02 | .00 | | .07 | .07 | -.01 | .00 | .04 |
| 10 | -.03 | -.02 | -.01 | .05 | -.03 | -.05 | .00 | -.01 | .08 | | .06 | -.02 | -.03 | -.02 |
| 12 | .01 | .05 | .02 | .05 | .04 | .00 | .05 | -.01 | .03 | .02 | | .08 | -.04 | -.02 |
| 14 | -.04 | -.05 | -.09 | -.04 | .01 | -.01 | .00 | -.03 | -.02 | -.03 | .05 | | -.05 | -.04 |
| 15 | -.01 | .02 | .00 | .06 | -.02 | -.04 | -.01 | -.03 | -.04 | -.02 | -.05 | -.02 | | .07 |
| 16 | .01 | .05 | .01 | .06 | .04 | .03 | .03 | .03 | .04 | .01 | -.01 | -.03 | .06 | |

*Note.* Sample 3 below the main diagonal, Sample 4 above the main diagonal. Residuals larger than |.10| are in boldface.

## Appendix D

Table D

*Study 2 Item Correlations and Descriptive Statistics*

| Item | $1_{T1}$ | $2_{T1}$ | $3_{T1}$ | $4_{T1}$ | $5_{T1}$ | $6_{T1}$ | $7_{T1}$ | $8_{T1}$ | $9_{T1}$ | $10_{T1}$ | $12_{T1}$ | $14_{T1}$ | $15_{T1}$ | $16_{T1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1_{T1}$ | 1. | | | | | | | | | | | | | |
| $2_{T1}$ | .59 | 1. | | | | | | | | | | | | |
| $3_{T1}$ | .71 | .69 | 1. | | | | | | | | | | | |
| $4_{T1}$ | .45 | .64 | .61 | 1. | | | | | | | | | | |
| $5_{T1}$ | .34 | .37 | .34 | .33 | 1. | | | | | | | | | |
| $6_{T1}$ | .45 | .45 | .43 | .37 | .63 | 1. | | | | | | | | |
| $7_{T1}$ | .36 | .42 | .41 | .37 | .59 | .67 | 1. | | | | | | | |
| $8_{T1}$ | .30 | .38 | .39 | .40 | .46 | .53 | .49 | 1. | | | | | | |
| $9_{T1}$ | .30 | .37 | .34 | .41 | .57 | .60 | .62 | .50 | 1. | | | | | |
| $10_{T1}$ | .28 | .38 | .36 | .43 | .53 | .56 | .58 | .50 | .71 | 1. | | | | |
| $12_{T1}$ | -.25 | -.24 | -.26 | -.19 | -.13 | -.18 | -.16 | -.13 | -.13 | -.17 | 1. | | | |
| $14_{T1}$ | -.37 | -.39 | -.45 | -.36 | -.25 | -.26 | -.30 | -.26 | -.25 | -.27 | .48 | 1. | | |
| $15_{T1}$ | -.26 | -.34 | -.32 | -.29 | -.21 | -.20 | -.22 | -.18 | -.23 | -.26 | .41 | .54 | 1. | |
| $16_{T1}$ | -.28 | -.28 | -.30 | -.24 | -.17 | -.18 | -.20 | -.13 | -.19 | -.22 | .42 | .49 | .62 | 1. |
| $1_{T2}$ | .28 | .22 | .26 | .16 | .15 | .21 | .15 | .14 | .15 | .18 | -.12 | -.21 | -.18 | -.17 |
| $2_{T2}$ | .20 | .28 | .26 | .27 | .24 | .24 | .22 | .15 | .22 | .23 | -.11 | -.24 | -.20 | -.16 |
| $3_{T2}$ | .23 | .24 | .29 | .18 | .17 | .23 | .18 | .15 | .20 | .20 | -.08 | -.21 | -.18 | -.14 |
| $4_{T2}$ | .19 | .27 | .28 | .27 | .20 | .23 | .23 | .18 | .21 | .20 | -.11 | -.24 | -.21 | -.18 |
| $5_{T2}$ | .15 | .16 | .15 | .16 | .24 | .24 | .21 | .20 | .21 | .21 | -.11 | -.16 | -.18 | -.15 |
| $6_{T2}$ | .15 | .16 | .16 | .17 | .20 | .24 | .20 | .18 | .22 | .22 | -.13 | -.15 | -.14 | -.11 |
| $7_{T2}$ | .11 | .17 | .14 | .17 | .19 | .20 | .21 | .16 | .20 | .21 | -.13 | -.15 | -.16 | -.11 |
| $8_{T2}$ | .13 | .16 | .17 | .17 | .17 | .22 | .21 | .23 | .20 | .22 | -.14 | -.17 | -.18 | -.14 |
| $9_{T2}$ | .10 | .17 | .13 | .14 | .16 | .20 | .20 | .18 | .19 | .20 | -.10 | -.15 | -.14 | -.09 |
| $10_{T2}$ | .12 | .18 | .13 | .20 | .20 | .24 | .21 | .20 | .25 | .27 | -.13 | -.17 | -.17 | -.11 |
| $12_{T2}$ | -.13 | -.13 | -.16 | -.12 | -.06 | -.14 | -.14 | -.10 | -.09 | -.11 | .33 | .28 | .24 | .20 |
| $14_{T2}$ | -.20 | -.22 | -.25 | -.14 | -.14 | -.19 | -.18 | -.12 | -.15 | -.15 | .25 | .35 | .24 | .24 |
| $15_{T2}$ | -.16 | -.18 | -.19 | -.14 | -.15 | -.19 | -.18 | -.16 | -.16 | -.18 | .21 | .28 | .31 | .25 |
| $16_{T2}$ | -.18 | -.21 | -.22 | -.16 | -.12 | -.18 | -.20 | -.13 | -.15 | -.19 | .24 | .31 | .32 | .32 |
| *SD* | 0.90 | 0.93 | 0.97 | 0.98 | 1.01 | 0.94 | 0.99 | 0.98 | 1.04 | 1.08 | 1.46 | 1.08 | 1.16 | 1.32 |
| Mean | 6.80 | 6.62 | 6.53 | 6.28 | 6.78 | 6.65 | 6.62 | 6.37 | 6.63 | 6.56 | 3.50 | 2.43 | 2.27 | 2.63 |
| Skew | -0.18 | -0.09 | -0.10 | 0.01 | -0.31 | -0.13 | -0.29 | -0.07 | -0.45 | -0.36 | 0.46 | 0.56 | 0.79 | 0.74 |
| Kurt | -0.83 | -0.65 | -0.63 | -0.25 | -0.57 | -0.71 | -0.31 | -0.16 | 0.18 | -0.20 | 0.14 | 0.29 | 0.51 | 0.41 |

*Study 2 Item Correlations and Descriptive Statistics, continued*

| Item | $1_{T2}$ | $2_{T2}$ | $3_{T2}$ | $4_{T2}$ | $5_{T2}$ | $6_{T2}$ | $7_{T2}$ | $8_{T2}$ | $9_{T2}$ | $10_{T2}$ | $12_{T2}$ | $14_{T2}$ | $15_{T2}$ | $16_{T2}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1_{T2}$ | 1. | | | | | | | | | | | | | |
| $2_{T2}$ | .66 | 1. | | | | | | | | | | | | |
| $3_{T2}$ | .75 | .74 | 1. | | | | | | | | | | | |
| $4_{T2}$ | .61 | .75 | .70 | 1. | | | | | | | | | | |
| $5_{T2}$ | .41 | .48 | .44 | .44 | 1. | | | | | | | | | |
| $6_{T2}$ | .47 | .52 | .47 | .47 | .80 | 1. | | | | | | | | |
| $7_{T2}$ | .41 | .50 | .44 | .46 | .72 | .79 | 1. | | | | | | | |
| $8_{T2}$ | .43 | .51 | .49 | .52 | .67 | .68 | .68 | 1. | | | | | | |
| $9_{T2}$ | .39 | .51 | .45 | .50 | .69 | .73 | .77 | .72 | 1. | | | | | |
| $10_{T2}$ | .42 | .53 | .48 | .55 | .67 | .72 | .76 | .69 | .78 | 1. | | | | |
| $12_{T2}$ | -.27 | -.24 | -.26 | -.24 | -.17 | -.21 | -.19 | -.20 | -.18 | -.16 | 1. | | | |
| $14_{T2}$ | -.43 | -.41 | -.44 | -.40 | -.29 | -.33 | -.32 | -.31 | -.32 | -.29 | .56 | 1. | | |
| $15_{T2}$ | -.29 | -.29 | -.30 | -.31 | -.26 | -.28 | -.30 | -.27 | -.29 | -.28 | .49 | .56 | 1. | |
| $16_{T2}$ | -.30 | -.28 | -.29 | -.32 | -.28 | -.28 | -.29 | -.27 | -.28 | -.26 | .55 | .61 | .72 | 1. |
| | | | | | | | | | | | | | | |
| *SD* | 0.97 | 0.96 | 0.99 | 0.99 | 1.12 | 1.10 | 1.13 | 1.18 | 1.14 | 1.17 | 1.59 | 1.21 | 1.36 | 1.28 |
| Mean | 6.55 | 6.59 | 6.44 | 6.47 | 6.58 | 6.40 | 6.43 | 6.22 | 6.48 | 6.36 | 3.90 | 2.88 | 2.84 | 2.98 |
| Skew | -0.42 | -0.46 | -0.38 | -0.36 | -0.96 | -0.65 | -0.74 | -0.51 | -0.71 | -0.61 | 0.43 | 0.84 | 0.78 | 0.58 |
| Kurt | 0.64 | 0.45 | 0.46 | 0.32 | 1.75 | 0.82 | 1.09 | 0.49 | 0.75 | 0.38 | -0.25 | 1.58 | 0.90 | 0.41 |

*Note*. $N$ = 912.

# Appendix E

Study 2 Correlation Residual Matrices

Table E1

*Correlation Residuals for the 14-Item Three-Factor Model*

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | -.03 | .06 | -.04 | -.02 | .02 | -.04 | .01 | -.06 | -.02 | -.03 | **-.15** | .01 | .02 |
| 2 | -.03 | | -.01 | .03 | .01 | .02 | .00 | .05 | .02 | .04 | .03 | **-.11** | .04 | .07 |
| 3 | .05 | -.02 | | -.02 | -.04 | -.03 | -.06 | .04 | -.05 | -.01 | .01 | **-.14** | .03 | .06 |
| 4 | -.09 | .06 | -.01 | | -.01 | .00 | -.01 | .09 | .03 | .08 | .02 | **-.11** | .01 | .02 |
| 5 | .00 | .01 | -.05 | .01 | | .07 | -.01 | .00 | -.03 | -.04 | .05 | -.04 | .01 | .01 |
| 6 | .08 | .06 | .01 | .02 | .04 | | .01 | -.02 | -.03 | -.03 | .02 | -.07 | .01 | .02 |
| 7 | -.01 | .03 | -.01 | .03 | .01 | .04 | | -.02 | .01 | .01 | .04 | -.05 | -.01 | .01 |
| 8 | .01 | .06 | .05 | **.12** | -.01 | .01 | -.02 | | .02 | .01 | .02 | -.06 | -.01 | .01 |
| 9 | -.07 | -.02 | -.09 | .06 | -.01 | -.04 | -.01 | -.01 | | .04 | .05 | -.06 | -.01 | .02 |
| 10 | -.07 | .00 | -.04 | .10 | -.03 | -.06 | -.03 | .01 | .10 | | .06 | -.04 | -.01 | .03 |
| 12 | -.01 | .01 | .02 | .04 | .03 | .00 | .02 | .01 | .05 | .00 | | .07 | -.04 | -.01 |
| 14 | -.07 | -.07 | -.10 | -.07 | -.04 | -.04 | -.07 | -.08 | -.02 | -.05 | .06 | | -.03 | -.03 |
| 15 | .06 | .00 | .04 | .00 | .00 | .03 | .02 | .01 | .01 | -.04 | -.04 | -.03 | | .04 |
| 16 | .02 | .05 | .05 | .04 | .04 | .05 | .03 | .05 | .03 | .00 | .00 | -.05 | .06 | |

*Note.* Time point 1 below the main diagonal, time point 2 above the main diagonal. Residuals larger than |.10| are in boldface.

Table E2

*Correlation Residuals for the Configural Invariance Model*

| Item | $1_{T1}$ | $2_{T1}$ | $3_{T1}$ | $4_{T1}$ | $5_{T1}$ | $6_{T1}$ | $7_{T1}$ | $8_{T1}$ | $9_{T1}$ | $10_{T1}$ | $12_{T1}$ | $14_{T1}$ | $15_{T1}$ | $16_{T1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1_{T1}$ | | | | | | | | | | | | | | |
| $2_{T1}$ | -.02 | | | | | | | | | | | | | |
| $3_{T1}$ | .06 | -.02 | | | | | | | | | | | | |
| $4_{T1}$ | -.09 | .06 | -.01 | | | | | | | | | | | |
| $5_{T1}$ | .01 | .01 | -.05 | .01 | | | | | | | | | | |
| $6_{T1}$ | .08 | .05 | .01 | .02 | .05 | | | | | | | | | |
| $7_{T1}$ | -.01 | .03 | -.01 | .03 | .01 | .04 | | | | | | | | |
| $8_{T1}$ | .01 | .06 | .05 | **.12** | -.01 | .01 | -.02 | | | | | | | |
| $9_{T1}$ | -.07 | -.02 | -.09 | .06 | -.01 | -.04 | -.01 | -.01 | | | | | | |
| $10_{T1}$ | -.07 | .00 | -.04 | .10 | -.03 | -.05 | -.03 | .01 | **.11** | | | | | |
| $12_{T1}$ | -.01 | .01 | .02 | .04 | .03 | .00 | .02 | .01 | .05 | .00 | | | | |
| $14_{T1}$ | -.07 | -.07 | -.10 | -.07 | -.04 | -.04 | -.07 | -.08 | -.02 | -.05 | .06 | | | |
| $15_{T1}$ | .06 | .00 | .04 | .01 | .00 | .03 | .02 | .01 | .01 | -.04 | -.04 | -.03 | | |
| $16_{T1}$ | .02 | .05 | .05 | .04 | .04 | .05 | .03 | .05 | .03 | -.01 | .00 | -.04 | .06 | |
| $1_{T2}$ | -.02 | -.01 | .02 | -.03 | -.03 | .01 | -.04 | -.02 | -.05 | .00 | .01 | -.04 | .00 | -.01 |
| $2_{T2}$ | -.03 | .02 | .00 | .05 | .05 | .02 | .01 | -.02 | .00 | .03 | .04 | -.05 | .00 | .03 |
| $3_{T2}$ | .00 | -.01 | -.02 | -.04 | -.03 | .01 | -.03 | -.02 | -.02 | .00 | .06 | -.03 | .01 | .04 |
| $4_{T2}$ | -.03 | .04 | .02 | .02 | .01 | .02 | .03 | .02 | .01 | .01 | .03 | -.06 | -.02 | .00 |
| $5_{T2}$ | .01 | .01 | -.01 | .03 | .00 | .04 | .00 | .03 | .00 | .01 | .00 | -.02 | -.03 | -.01 |
| $6_{T2}$ | .00 | .01 | -.01 | .03 | .00 | .00 | -.02 | .01 | .00 | .01 | -.01 | .00 | .01 | .03 |
| $7_{T2}$ | -.03 | .01 | -.03 | .03 | -.01 | -.02 | -.03 | -.02 | -.02 | .00 | -.01 | .00 | .00 | .03 |
| $8_{T2}$ | .00 | .02 | .02 | .05 | -.01 | .01 | .01 | .02 | .00 | .03 | -.04 | -.04 | -.04 | .00 |
| $9_{T2}$ | -.04 | .01 | -.04 | .01 | -.04 | -.02 | -.02 | .01 | -.02 | -.01 | .01 | .00 | .01 | .06 |
| $10_{T2}$ | -.02 | .03 | -.03 | .06 | .01 | .03 | .00 | .03 | .04 | .03 | -.02 | -.03 | -.02 | .03 |
| $12_{T2}$ | .01 | .02 | .01 | .02 | .06 | .00 | .00 | .01 | .05 | .02 | .01 | .05 | .00 | -.03 |
| $14_{T2}$ | -.04 | -.05 | -.06 | .01 | .00 | -.03 | -.03 | .00 | .00 | .00 | .05 | .02 | -.03 | -.01 |
| $15_{T2}$ | .01 | .01 | .02 | .03 | .00 | -.02 | -.01 | -.02 | .01 | -.02 | -.01 | .00 | -.01 | -.02 |
| $16_{T2}$ | .01 | .00 | .00 | .02 | .04 | .00 | -.02 | .01 | .03 | -.02 | .00 | .02 | .01 | -.01 |

*Correlation Residuals for the Configural Invariance Model, continued*

| Item | $1_{T2}$ | $2_{T2}$ | $3_{T2}$ | $4_{T2}$ | $5_{T2}$ | $6_{T2}$ | $7_{T2}$ | $8_{T2}$ | $9_{T2}$ | $10_{T2}$ | $12_{T2}$ | $14_{T2}$ | $15_{T2}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1_{T2}$ | | | | | | | | | | | | | |
| $2_{T2}$ | -.03 | | | | | | | | | | | | |
| $3_{T2}$ | .06 | -.01 | | | | | | | | | | | |
| $4_{T2}$ | -.04 | .03 | -.02 | | | | | | | | | | |
| $5_{T2}$ | -.02 | .01 | -.03 | -.01 | | | | | | | | | |
| $6_{T2}$ | .02 | .02 | -.03 | .00 | .07 | | | | | | | | |
| $7_{T2}$ | -.04 | .00 | -.06 | -.02 | -.01 | .01 | | | | | | | |
| $8_{T2}$ | .01 | .05 | .04 | .08 | .00 | -.02 | -.02 | | | | | | |
| $9_{T2}$ | -.06 | .02 | -.05 | .03 | -.03 | -.03 | .01 | .03 | | | | | |
| $10_{T2}$ | -.02 | .04 | -.01 | .08 | -.04 | -.03 | .01 | .01 | .04 | | | | |
| $12_{T2}$ | -.03 | .03 | .01 | .02 | .05 | .02 | .04 | .01 | .05 | .06 | | | |
| $14_{T2}$ | **-.16** | **-.11** | **-.14** | **-.11** | -.04 | -.07 | -.05 | -.07 | -.06 | -.04 | .08 | | |
| $15_{T2}$ | .01 | .04 | .03 | .01 | .01 | .01 | -.01 | -.01 | -.01 | -.01 | -.04 | -.03 | |
| $16_{T2}$ | .02 | .07 | .06 | .02 | .01 | .02 | .01 | .01 | .02 | .03 | -.01 | -.03 | .03 |

Table E3

*Correlation Residuals for the Metric Invariance Model*

| Item | $1_{T1}$ | $2_{T1}$ | $3_{T1}$ | $4_{T1}$ | $5_{T1}$ | $6_{T1}$ | $7_{T1}$ | $8_{T1}$ | $9_{T1}$ | $10_{T1}$ | $12_{T1}$ | $14_{T1}$ | $15_{T1}$ | $16_{T1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1_{T1}$ | | | | | | | | | | | | | | |
| $2_{T1}$ | -.04 | | | | | | | | | | | | | |
| $3_{T1}$ | .06 | -.01 | | | | | | | | | | | | |
| $4_{T1}$ | -.10 | .04 | -.02 | | | | | | | | | | | |
| $5_{T1}$ | .00 | .00 | -.05 | .00 | | | | | | | | | | |
| $6_{T1}$ | .07 | .04 | .01 | .00 | .04 | | | | | | | | | |
| $7_{T1}$ | -.02 | .02 | -.01 | .01 | .01 | .03 | | | | | | | | |
| $8_{T1}$ | -.02 | .03 | .03 | .08 | -.04 | -.03 | -.06 | | | | | | | |
| $9_{T1}$ | -.07 | -.02 | -.07 | .05 | .00 | -.03 | .00 | -.03 | | | | | | |
| $10_{T1}$ | -.07 | .00 | -.03 | .09 | -.02 | -.05 | -.02 | -.02 | **.13** | | | | | |
| $12_{T1}$ | -.01 | .02 | .02 | .05 | .04 | .00 | .02 | .02 | .05 | .00 | | | | |
| $14_{T1}$ | -.07 | -.07 | **-.12** | -.07 | -.05 | -.04 | -.07 | -.07 | -.03 | -.06 | .05 | | | |
| $15_{T1}$ | .06 | .00 | .04 | .02 | .00 | .04 | .02 | .02 | .01 | -.04 | -.05 | -.03 | | |
| $16_{T1}$ | .02 | .05 | .04 | .05 | .04 | .04 | .02 | .06 | .03 | -.01 | .00 | -.03 | .06 | |
| $1_{T2}$ | -.02 | -.01 | .02 | -.04 | -.02 | .01 | -.04 | -.03 | -.04 | .00 | .02 | -.04 | .00 | -.01 |
| $2_{T2}$ | -.03 | .02 | .00 | .05 | .05 | .02 | .01 | -.03 | .01 | .03 | .04 | -.06 | .00 | .02 |
| $3_{T2}$ | .00 | -.01 | -.01 | -.05 | -.03 | .01 | -.04 | -.04 | -.01 | .00 | .07 | -.03 | .01 | .04 |
| $4_{T2}$ | -.03 | .04 | .03 | .02 | .01 | .03 | .03 | .01 | .01 | .01 | .03 | -.07 | -.02 | -.01 |
| $5_{T2}$ | .01 | .01 | -.01 | .03 | .00 | .03 | .01 | .02 | .00 | .02 | .00 | -.02 | -.03 | -.01 |
| $6_{T2}$ | .00 | .01 | -.01 | .02 | .00 | .00 | -.02 | -.01 | .01 | .02 | -.01 | .00 | .01 | .03 |
| $7_{T2}$ | -.04 | .01 | -.02 | .02 | -.01 | -.03 | -.03 | -.03 | -.02 | .00 | -.01 | -.01 | .00 | .03 |
| $8_{T2}$ | .00 | .02 | .02 | .04 | .00 | .02 | .02 | .01 | .01 | .04 | -.04 | -.04 | -.04 | -.01 |
| $9_{T2}$ | -.05 | .01 | -.04 | .00 | -.04 | -.02 | -.02 | -.01 | -.01 | -.01 | .02 | .00 | .01 | .06 |
| $10_{T2}$ | -.03 | .03 | -.03 | .06 | .00 | .03 | .00 | .02 | .04 | .03 | -.01 | -.03 | -.02 | .03 |
| $12_{T2}$ | .01 | .02 | .00 | .02 | .06 | .00 | -.01 | .02 | .05 | .02 | .01 | .06 | .00 | -.02 |
| $14_{T2}$ | -.03 | -.05 | -.06 | .02 | .00 | -.03 | -.02 | .01 | .00 | .00 | .04 | .02 | -.03 | -.01 |
| $15_{T2}$ | .01 | .01 | .01 | .03 | .00 | -.03 | -.01 | -.01 | .00 | -.03 | -.01 | .01 | -.01 | -.02 |
| $16_{T2}$ | .01 | .00 | .00 | .03 | .04 | .00 | -.02 | .02 | .02 | -.02 | .00 | .02 | .00 | -.01 |

*Correlation Residuals for the Metric Invariance Model, continued*

| Item | $1_{T2}$ | $2_{T2}$ | $3_{T2}$ | $4_{T2}$ | $5_{T2}$ | $6_{T2}$ | $7_{T2}$ | $8_{T2}$ | $9_{T2}$ | $10_{T2}$ | $12_{T2}$ | $14_{T2}$ | $15_{T2}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1_{T2}$ | | | | | | | | | | | | | |
| $2_{T2}$ | -.02 | | | | | | | | | | | | |
| $3_{T2}$ | .06 | -.02 | | | | | | | | | | | |
| $4_{T2}$ | -.03 | .04 | -.02 | | | | | | | | | | |
| $5_{T2}$ | -.01 | .01 | -.04 | -.01 | | | | | | | | | |
| $6_{T2}$ | .02 | .02 | -.03 | .00 | .07 | | | | | | | | |
| $7_{T2}$ | -.04 | .00 | -.06 | -.01 | -.01 | .01 | | | | | | | |
| $8_{T2}$ | .02 | .07 | .05 | .10 | .02 | .00 | -.01 | | | | | | |
| $9_{T2}$ | -.06 | .02 | -.05 | .03 | -.04 | -.04 | .00 | .04 | | | | | |
| $10_{T2}$ | -.02 | .04 | -.01 | .09 | -.05 | -.03 | .00 | .02 | .03 | | | | |
| $12_{T2}$ | -.03 | .02 | .01 | .01 | .05 | .01 | .04 | .01 | .05 | .06 | | | |
| $14_{T2}$ | **-.15** | **-.11** | **-.13** | **-.11** | -.04 | -.07 | -.05 | -.07 | -.06 | -.03 | .08 | | |
| $15_{T2}$ | .00 | .04 | .03 | .00 | .01 | .00 | -.01 | -.02 | -.01 | -.01 | -.03 | -.03 | |
| $16_{T2}$ | .02 | .07 | .06 | .02 | .01 | .02 | .02 | .00 | .02 | .03 | .00 | -.03 | .04 |

Table E4

*Correlation Residuals for the Scalar Invariance Model*

| Item | $1_{T1}$ | $2_{T1}$ | $3_{T1}$ | $4_{T1}$ | $5_{T1}$ | $6_{T1}$ | $7_{T1}$ | $8_{T1}$ | $9_{T1}$ | $10_{T1}$ | $12_{T1}$ | $14_{T1}$ | $15_{T1}$ | $16_{T1}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1_{T1}$ | | | | | | | | | | | | | | |
| $2_{T1}$ | -.03 | | | | | | | | | | | | | |
| $3_{T1}$ | .07 | -.01 | | | | | | | | | | | | |
| $4_{T1}$ | -.09 | .05 | -.01 | | | | | | | | | | | |
| $5_{T1}$ | .00 | .00 | -.05 | .01 | | | | | | | | | | |
| $6_{T1}$ | .07 | .04 | .01 | .01 | .04 | | | | | | | | | |
| $7_{T1}$ | -.01 | .03 | -.01 | .02 | .01 | .03 | | | | | | | | |
| $8_{T1}$ | -.02 | .03 | .03 | .09 | -.04 | -.03 | -.06 | | | | | | | |
| $9_{T1}$ | -.06 | -.02 | -.07 | .06 | .00 | -.03 | .00 | -.03 | | | | | | |
| $10_{T1}$ | -.07 | .00 | -.03 | .10 | -.02 | -.05 | -.01 | -.02 | **.13** | | | | | |
| $12_{T1}$ | -.01 | .02 | .02 | .05 | .04 | .00 | .02 | .02 | .05 | .00 | | | | |
| $14_{T1}$ | -.07 | -.06 | **-.11** | -.07 | -.04 | -.04 | -.07 | -.07 | -.03 | -.06 | .05 | | | |
| $15_{T1}$ | .06 | .01 | .04 | .01 | .01 | .04 | .02 | .02 | .01 | -.04 | -.05 | -.03 | | |
| $16_{T1}$ | .01 | .04 | .04 | .04 | .03 | .04 | .02 | .05 | .02 | -.02 | .01 | -.02 | .07 | |
| $1_{T2}$ | -.01 | -.01 | .02 | -.04 | -.02 | .01 | -.04 | -.03 | -.04 | .00 | .01 | -.04 | .00 | -.02 |
| $2_{T2}$ | -.03 | .02 | .00 | .05 | .05 | .02 | .01 | -.03 | .01 | .03 | .04 | -.06 | .00 | .02 |
| $3_{T2}$ | .00 | -.01 | -.02 | -.05 | -.03 | .01 | -.04 | -.04 | -.01 | .00 | .07 | -.03 | .02 | .03 |
| $4_{T2}$ | -.03 | .04 | .03 | .04 | .01 | .03 | .03 | .01 | .02 | .02 | .03 | -.07 | -.02 | -.01 |
| $5_{T2}$ | .01 | .01 | -.01 | .03 | .00 | .03 | .01 | .02 | .00 | .02 | .00 | -.02 | -.03 | -.01 |
| $6_{T2}$ | .00 | .01 | -.01 | .03 | .00 | .00 | -.02 | -.01 | .01 | .02 | -.01 | .00 | .01 | .03 |
| $7_{T2}$ | -.04 | .01 | -.02 | .03 | -.01 | -.03 | -.03 | -.03 | -.02 | .00 | -.01 | .00 | .00 | .03 |
| $8_{T2}$ | .00 | .02 | .02 | .05 | .00 | .02 | .02 | .01 | .01 | .04 | -.04 | -.04 | -.04 | -.01 |
| $9_{T2}$ | -.05 | .01 | -.04 | .00 | -.04 | -.02 | -.02 | -.01 | -.01 | -.01 | .02 | .00 | .01 | .05 |
| $10_{T2}$ | -.03 | .03 | -.03 | .06 | .00 | .02 | .00 | .02 | .04 | .03 | -.01 | -.03 | -.02 | .02 |
| $12_{T2}$ | .01 | .02 | .00 | .02 | .06 | .00 | -.01 | .02 | .05 | .02 | .01 | .05 | .00 | -.02 |
| $14_{T2}$ | -.03 | -.04 | -.06 | .02 | .00 | -.03 | -.02 | .01 | .00 | .00 | .04 | .02 | -.04 | -.01 |
| $15_{T2}$ | .01 | .01 | .01 | .03 | .00 | -.02 | -.01 | -.01 | .00 | -.03 | -.02 | .00 | -.02 | -.01 |
| $16_{T2}$ | .01 | .00 | .00 | .02 | .04 | .00 | -.02 | .02 | .02 | -.02 | .00 | .02 | .00 | .00 |

*Correlation Residuals for the Scalar Invariance Model, continued*

| Item | $1_{T2}$ | $2_{T2}$ | $3_{T2}$ | $4_{T2}$ | $5_{T2}$ | $6_{T2}$ | $7_{T2}$ | $8_{T2}$ | $9_{T2}$ | $10_{T2}$ | $12_{T2}$ | $14_{T2}$ | $15_{T2}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1_{T2}$ | | | | | | | | | | | | | |
| $2_{T2}$ | -.02 | | | | | | | | | | | | |
| $3_{T2}$ | .06 | -.02 | | | | | | | | | | | |
| $4_{T2}$ | -.02 | .05 | -.01 | | | | | | | | | | |
| $5_{T2}$ | -.01 | .01 | -.04 | .00 | | | | | | | | | |
| $6_{T2}$ | .02 | .02 | -.03 | .01 | .07 | | | | | | | | |
| $7_{T2}$ | -.04 | .00 | -.06 | -.01 | -.01 | .01 | | | | | | | |
| $8_{T2}$ | .03 | .07 | .05 | .10 | .02 | .00 | -.01 | | | | | | |
| $9_{T2}$ | -.06 | .02 | -.05 | .04 | -.03 | -.04 | .01 | .04 | | | | | |
| $10_{T2}$ | -.02 | .04 | -.01 | .09 | -.05 | -.03 | .00 | .02 | .04 | | | | |
| $12_{T2}$ | -.03 | .02 | .01 | .01 | .05 | .02 | .04 | .01 | .05 | .06 | | | |
| $14_{T2}$ | **-.15** | -.10 | **-.13** | **-.11** | -.03 | -.07 | -.05 | -.07 | -.05 | -.03 | .08 | | |
| $15_{T2}$ | .01 | .04 | .04 | .00 | .01 | .01 | -.01 | -.02 | .00 | .00 | -.03 | -.04 | |
| $16_{T2}$ | .02 | .07 | .06 | .01 | .01 | .02 | .01 | .00 | .02 | .03 | .01 | -.03 | .04 |

Table 1

*Expectancy-Value Constructs and Example Items*

| Construct / subtype | Description and example items |
|---|---|
| Expectancy | An individual's belief about how well he or she will do on an upcoming task |
| *Ability beliefs* | *How good at math are you?* |
| *Expectancy* | *How well do you think you will do in your math course this year?* |
| Value | The reason(s) an individual engages in or attempts to succeed at a task |
| *Attainment value* | *For me, being good in math is (not at all – very important).* |
| *Intrinsic value* | *How much do you like doing math?* |
| *Utility value* | *In general, how useful is what you learn in math?* |
| Cost | The extent to which successfully engaging in an activity is constrained by other factors |
| | *I have to give up a lot to do well in math.* |

*Note*. Example expectancy and value items from Wigfield (1994); cost item from Conley (2012).

Table 2

*Initial Samples for Study 1 and Study 2*

| | | Both time points | | | |
|---|---|---|---|---|---|
| Sample | T1 data only | T1 data | T2 data | T2 data only | Total |
| Sample 1 | 359 | 113 | 228 | 244 | 944 |
| Sample 2 | 359 | 113 | 228 | 244 | 944 |
| Sample 3 | 359 | 113 | 227 | 245 | 944 |
| Sample 4 | 360 | 112 | 227 | 245 | 944 |
| Matched | 0 | 951 | | 0 | 951 |
| Total | 1,437 | 2,312 | | 978 | 4,727 |

*Note.* In Samples 1-4, a portion of cases—approximately 36 percent—had complete data for both time points (i.e., they were matched cases). However, data from only one time point was used in each case. For instance, in Sample 1, 341 cases (113 + 228) had complete data at both time points; T1 data were used for 113 of the cases, and T2 data were used for the other 228 cases. All sample sizes were reduced after the removal of multivariate outliers.

Table 3

*Study 1 Sample Characteristics after the Removal of Multivariate Outliers*

| Characteristic | Sample 1 (*N* = 921) | Sample 2 (*N* = 923) | Sample 3 (*N* = 916) | Sample 4 (*N* = 922) |
|---|---|---|---|---|
| T1/first-year students | 462 | 467 | 457 | 463 |
| T2/mid-career students | 459 | 456 | 459 | 459 |
| Computer-based format (%) | 10.97 | 9.21 | 8.52 | 9.44 |
| Female (%) | 59.17 | 58.46 | 60.04 | 58.52 |
| Race/ethnicity (%)[a] | | | | |
| American Indian | 1.25 | 0.71 | 1.10 | 1.42 |
| Asian | 6.06 | 6.61 | 7.55 | 6.55 |
| Black | 6.06 | 5.54 | 5.40 | 5.66 |
| Hispanic | 4.10 | 4.82 | 3.96 | 5.84 |
| Pacific | 0.89 | 0.89 | 0.72 | 0.35 |
| White | 88.06 | 87.86 | 88.13 | 88.14 |
| Mean age | 19.42 | 19.42 | 19.37 | 19.44 |
| (*SD*) | (1.48) | (1.75) | (1.34) | (1.67) |
| Mean SAT-Math score | 572.40 | 575.21 | 574.51 | 573.91 |
| (*SD*) | (68.17) | (69.65) | (67.24) | (73.13) |
| Mean SAT-Verbal score | 564.95 | 565.20 | 565.39 | 563.61 |
| (*SD*) | (74.44) | (72.54) | (74.62) | (70.30) |
| Mean cumulative GPA[b] | 2.81 | 2.78 | 2.78 | 2.73 |
| (*SD*) | (0.87) | (0.84) | (0.84) | (0.93) |
| Mean earned credits[b] | 55.79 | 56.15 | 55.97 | 55.53 |
| (*SD*) | (7.83) | (7.92) | (8.19) | (7.89) |

[a] Race/ethnicity reported only for students who specified their race/ethnicity at T2 (61 percent of each sample). Percentages may sum to more than 100% in each sample because students were able to select multiple responses. Race/ethnicity from students with T1 data only is not included in the figures.

[b] Mean cumulative GPA and earned credits reported only for the mid-career (T2) students in each sample (approximately half of each sample). Incoming (T1) students did not yet have a GPA or credits earned through the university.

Table 4

*Study 1 Model Fit Indices*

| Model | *df* | $\chi^2_{SB}$ | $RMSEA_{SB}$ | SRMR | $CFI_{SB}$ | AIC |
|---|---|---|---|---|---|---|
| Sample 1 | | | | | | |
| Unidimensional | 90 | 2313.28 | .20 | .14 | .76 | 3431.21 |
| Two-factor | 89 | 1660.09 | .16 | .13 | .84 | 2309.04 |
| Bifactor | 84 | 546.56 | .08 | **.07** | **.96** | 664.92 |
| Three-factor | 87 | 495.74 | .08 | **.04** | **.97** | 609.88 |
| Sample 2 | | | | | | |
| Unidimensional | 90 | 2128.00 | .19 | .13 | .79 | 3134.78 |
| Two-factor | 89 | 1506.35 | .15 | .12 | .87 | 2091.89 |
| Bifactor | 84 | 527.06 | .08 | **.06** | **.97** | 622.38 |
| Three-factor | 87 | 474.22 | .07 | **.04** | **.97** | 565.82 |
| Sample 3 | | | | | | |
| Unidimensional | 77 | 2115.67 | .20 | .13 | .75 | 3015.25 |
| Two-factor | 76 | 1317.38 | .15 | .12 | .86 | 1733.38 |
| Bifactor | 72 | 370.25 | .07 | **.06** | **.97** | 451.21 |
| Three-factor | 74 | 328.19 | .06 | **.03** | **.98** | 406.86 |
| Sample 4 | | | | | | |
| Unidimensional | 77 | 1767.63 | .18 | .12 | .82 | 2512.82 |
| Two-factor | 76 | 1093.33 | .14 | .11 | .90 | 1425.45 |
| Bifactor | 72 | 404.81 | .07 | **.06** | **.97** | 497.41 |
| Three-factor | 74 | 369.31 | .07 | **.04** | **.98** | 458.96 |

*Note.* Fit indices that meet recommended cutoffs are shown in boldface ($RMSEA_{SB} \leq .05$, $SRMR \leq .07$, $CFI_{SB} \geq .96$; Yu & Muthén, 2002). AIC does not have an absolute cutoff; rather, a model with a smaller AIC provides more parsimonious fit than one with a larger AIC. Item 11 was dropped from the analyses in Samples 3 and 4, hence, degrees of freedom differ across models of the same name.

Table 5

*Parameter and Reliability Estimates for the 14-Item Three-Factor Model*

| | Sample 3 | | Sample 4 | |
|---|---|---|---|---|
| Item | Unstandardized pattern coefficient (*SE*) | Error, $1-R^2$ | Unstandardized pattern coefficient (*SE*) | Error, $1-R^2$ |
| 1 | .73 (.03) | .40 | .69 (.03) | .45 |
| 2 | .82 (.03) | .29 | .80 (.02) | .31 |
| 3 | .89 (.03) | .26 | .87 (.03) | .23 |
| 4 | .88 (.03) | .34 | .80 (.03) | .44 |
| 5 | .82 (.03) | .38 | .83 (.03) | .37 |
| 6 | .93 (.03) | .25 | .88 (.03) | .29 |
| 7 | .94 (.03) | .28 | .93 (.03) | .30 |
| 8 | .92 (.04) | .40 | .81 (.03) | .48 |
| 9 | 1.02 (.04) | .28 | .99 (.04) | .26 |
| 10 | .96 (.04) | .34 | 1.00 (.03) | .30 |
| 12 | 1.17 (.05) | .51 | 1.03 (.06) | .60 |
| 14 | 1.11 (.05) | .33 | .93 (.04) | .46 |
| 15 | 1.11 (.05) | .40 | .98 (.05) | .47 |
| 16 | 1.18 (.05) | .38 | 1.08 (.04) | .41 |
| Expectancy ω: | .89 | | .88 | |
| Value ω: | .93 | | .92 | |
| Cost ω: | .85 | | .80 | |

*Note*. *SE* = standard error. To set the metric of the latent variables, the latent factor means and variances were fixed to zero and one, respectively. Standardized parameter estimates are shown in Figures 11 and 12.

Table 6

*Study 2 Sample Characteristics after the Removal of Multivariate Outliers*

| Characteristic | Matched sample (*N* = 912) |
|---|---|
| Computer-based format at T1 (%) | 8.00 |
| Computer-based format at T2 (%) | 8.55 |
| Female (%) | 64.91 |
| Race/ethnicity (%) | |
| American Indian | .66 |
| Asian | 6.25 |
| Black | 2.63 |
| Hispanic | 3.73 |
| Pacific | 2.19 |
| White | 89.36 |
| Mean age at T1 (*SD*) | 18.42 (.39) |
| Mean age at T2 (*SD*) | 19.89 (.39) |
| Mean SAT-Math score (*SD*) | 579.91 (64.69) |
| Mean SAT-Verbal score (*SD*) | 571.60 (69.52) |
| Mean cumulative GPA (*SD*)[a] | 3.12 (.43) |
| Mean earned credits (*SD*)[a] | 52.30 (6.26) |

[a] Mean cumulative GPA and earned credits reported as of time point 2 (mid-career).

Table 7

*Study 2 Model Fit Indices*

| Model | *df* | $\chi^2_{SB}$ | $RMSEA_{SB}$ | SRMR | $CFI_{SB}$ |
|---|---|---|---|---|---|
| T1 only | 74 | 385.40 | .07 | **.04** | **.97** |
| T2 only | 74 | 397.18 | .07 | **.04** | **.97** |
| Configural | 321 | 998.92 | **.05** | **.04** | **.98** |
| Metric | 332 | 1033.14 | **.05** | **.04** | **.98** |
| Scalar | 343 | 1203.69 | .06 | **.04** | **.97** |

*Note.* Fit indices that meet recommended cutoffs are shown in boldface ($RMSEA_{SB} \leq .05$, $SRMR \leq .07$, $CFI_{SB} \geq .96$; Yu & Muthén, 2002).

Table 8

*Parameter and Reliability Estimates for the Scalar Invariant Model*

| Item | Intercept | Unstandardized pattern coefficient | Standardized pattern coefficient (T1, T2) | Error, $1-R^2$ (T1, T2) | Autocorrelation |
|---|---|---|---|---|---|
| 1 | 1.12 | .86 | .76, .78 | .43, .39 | .08 |
| 2 | .55 | .93 | .82, .87 | .34, .25 | .01 |
| 3* | .00 | 1.00 | .86, .88 | .27, .23 | .04 |
| 4 | .60 | .90 | .72, .81 | .48, .35 | .03 |
| 5 | .37 | .97 | .73, .83 | .46, .31 | .05 |
| 6* | .00 | 1.00 | .81, .88 | .35, .23 | .02 |
| 7 | -.21 | 1.03 | .80, .88 | .37, .23 | .02 |
| 8 | .25 | .93 | .69, .78 | .53, .39 | .05 |
| 9 | -.25 | 1.05 | .78, .87 | .39, .24 | .00 |
| 10 | -.39 | 1.05 | .75, .86 | .44, .27 | .04 |
| 12 | 1.31 | .94 | .59, .65 | .65, .58 | .14 |
| 14 | .49 | .85 | .73, .75 | .47, .43 | .07 |
| 15* | .00 | 1.00 | .78, .80 | .39, .36 | .03 |
| 16 | .24 | 1.00 | .71, .85 | .50, .28 | .04 |
| Expectancy ω: .87, .90 | | | | | |
| Value ω: .89, .94 | | | | | |
| Cost ω: .79, .84 | | | | | |

*Note.* To set the metric of the latent variables, the unstandardized intercepts and pattern coefficients were fixed to zero and one, respectively, for referent indicator items 3 (Expectancy), 6 (Value), and 15 (Cost).

Table 9

*Latent and Observed Mean Differences*

| | Latent | | | | | Observed | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Factor | T1 *M* (*SD*) | T2 *M* (*SD*) | Δ | *p* | *d* | T1 *M* (*SD*) | T2 *M* (*SD*) | Δ | *p* | *d* |
| Expectancy | 6.51 (.82) | 6.45 (.89) | -.06 | .78 | -.06 | 6.55 (.80) | 6.51 (.86) | -.04 | .20 | -.04 |
| Value | 6.62 (.76) | 6.42 (.97) | -.20 | .37 | -.19 | 6.60 (.81) | 6.41 (1.00) | -.19 | <.01 | -.18 |
| Cost | 2.32 (.92) | 2.79 (1.09) | .47 | <.05 | .46 | 2.71 (.99) | 3.15 (1.12) | .44 | <.01 | .38 |

*Note*. *N* = 912. Means (*M*), standard deviations (*SD*), and mean differences (Δ) are reported on a 1-8 response scale. Mean differences are computed as T2 mean minus T1 mean. A negative mean difference indicates a decrease from T1 to T2; a positive mean difference indicates an increase from T1 to T2.

Table 10

*Latent and Observed Factor Correlations*

| Factor | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. $Expectancy_{T1}$ | | .62 | -.55 | .35 | .22 | -.29 |
| 2. $Value_{T1}$ | .55 | | -.39 | .31 | .31 | -.26 |
| 3. $Cost_{T1}$ | -.45 | -.32 | | -.29 | -.23 | .47 |
| 4. $Expectancy_{T2}$ | .33 | .28 | -.24 | | .65 | -.48 |
| 5. $Value_{T2}$ | .21 | .30 | -.20 | .61 | | -.41 |
| 6. $Cost_{T2}$ | -.25 | -.22 | .42 | -.43 | -.36 | |

*Note.* $N$ = 912. Observed correlations below the main diagonal, latent correlations above the main diagonal.

Table 11

*Individual Change from Time Point 1 to Time Point 2*

| | Decrease | | No change | | Increase | |
|---|---|---|---|---|---|---|
| Scale | *n* | % | *n* | % | *n* | % |
| Expectancy | 396 | 43.42 | 141 | 15.46 | 375 | 41.12 |
| Value | 461 | 50.55 | 70 | 7.68 | 381 | 41.78 |
| Cost | 282 | 30.92 | 89 | 9.76 | 541 | 59.32 |

*Note. N* = 912. Change calculated as total scale score at time point 2 minus total scale score at time point 1. Negative differences (< 0) are considered decreases, and positive differences (> 0) are considered increases.

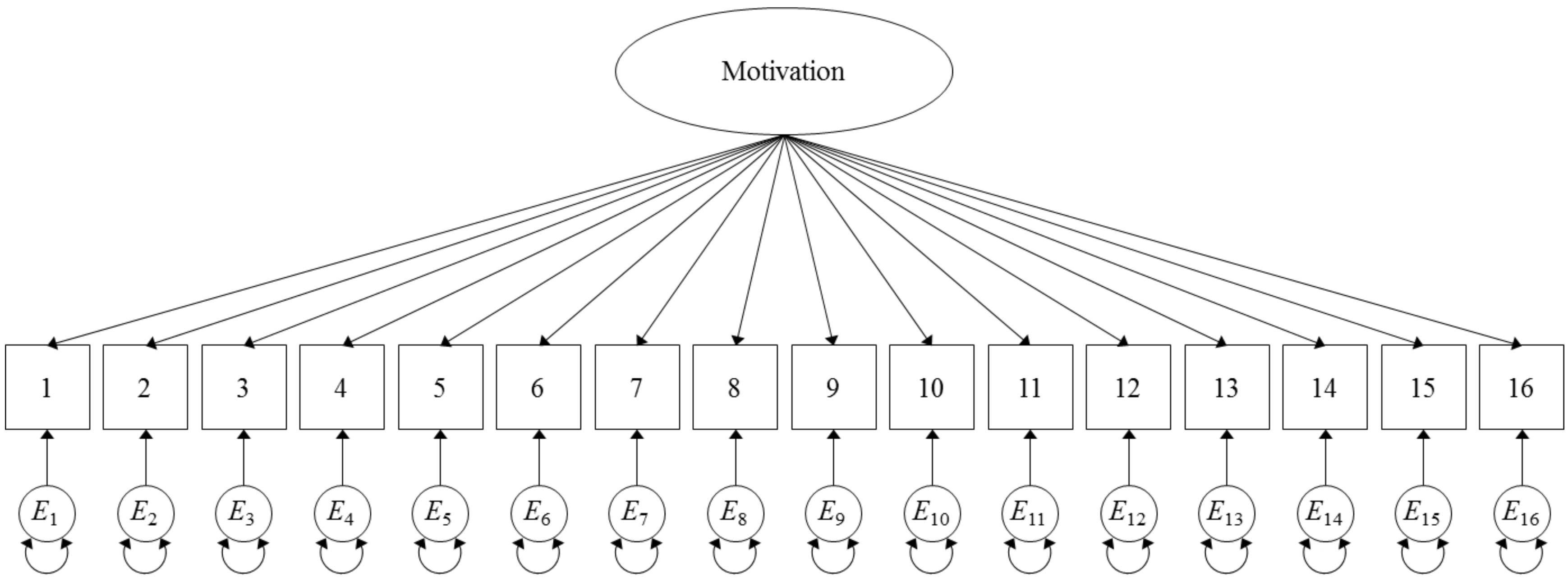


*Figure 1.* Unidimensional model.

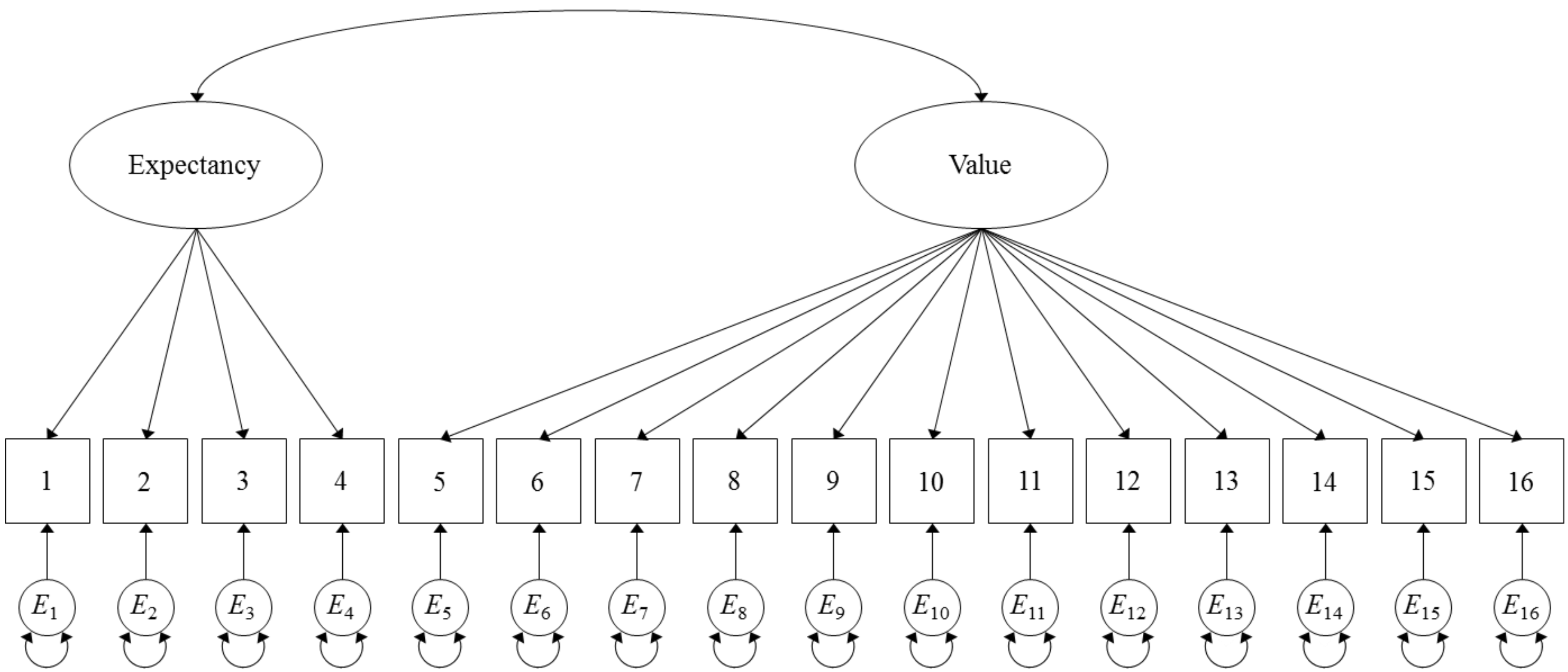


*Figure 2.* Correlated two-factor model.

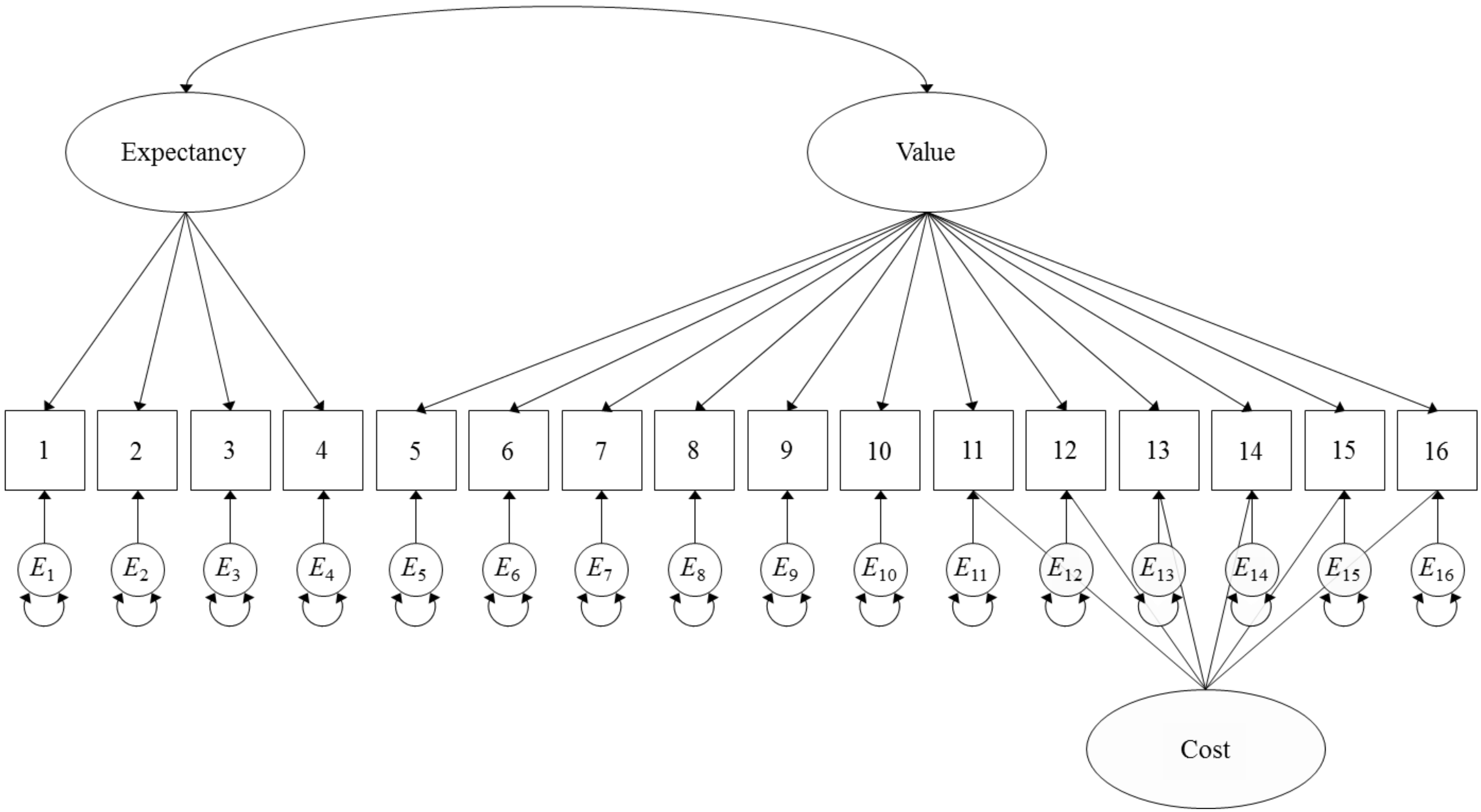


*Figure 3.* Bifactor model.

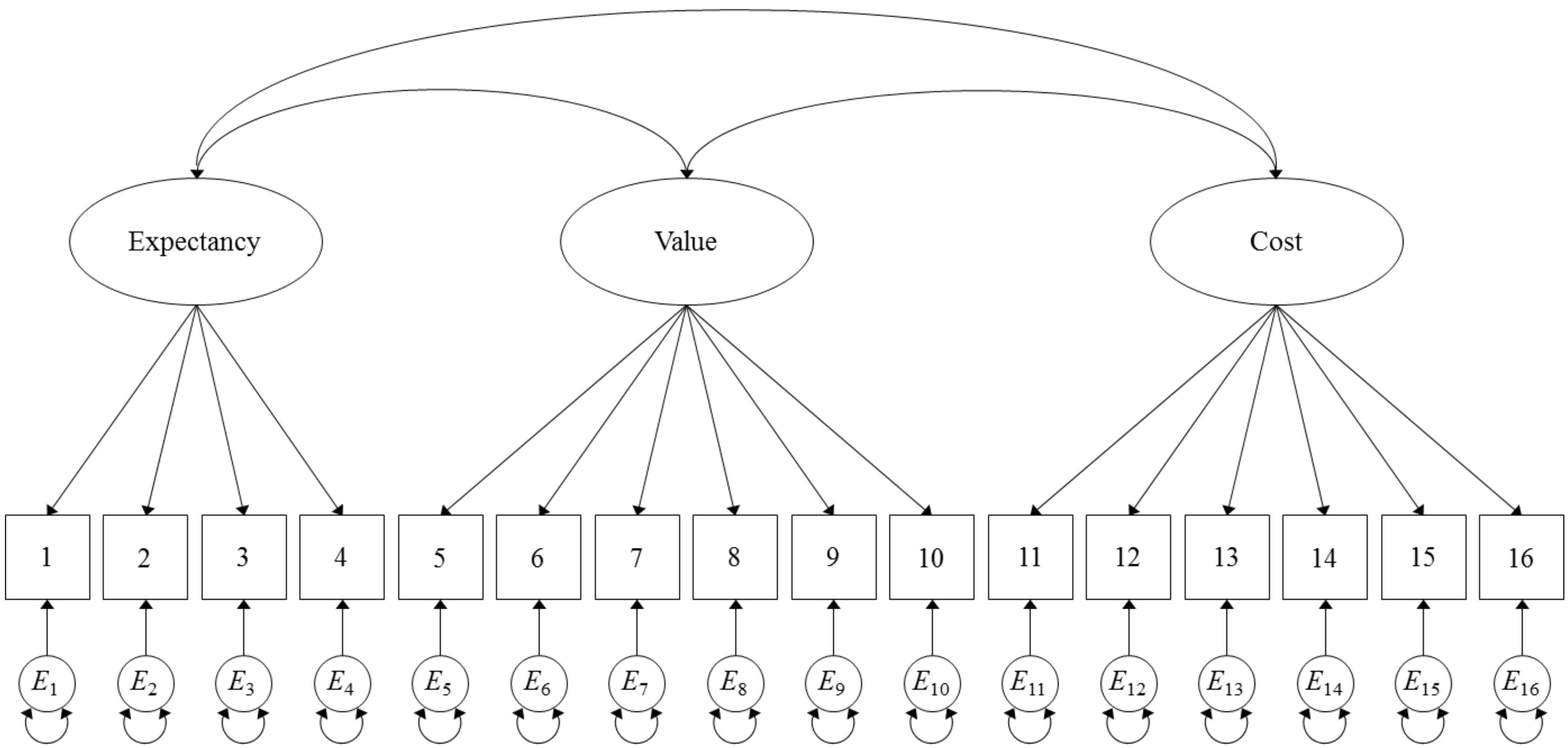


*Figure 4.* Correlated three-factor model.

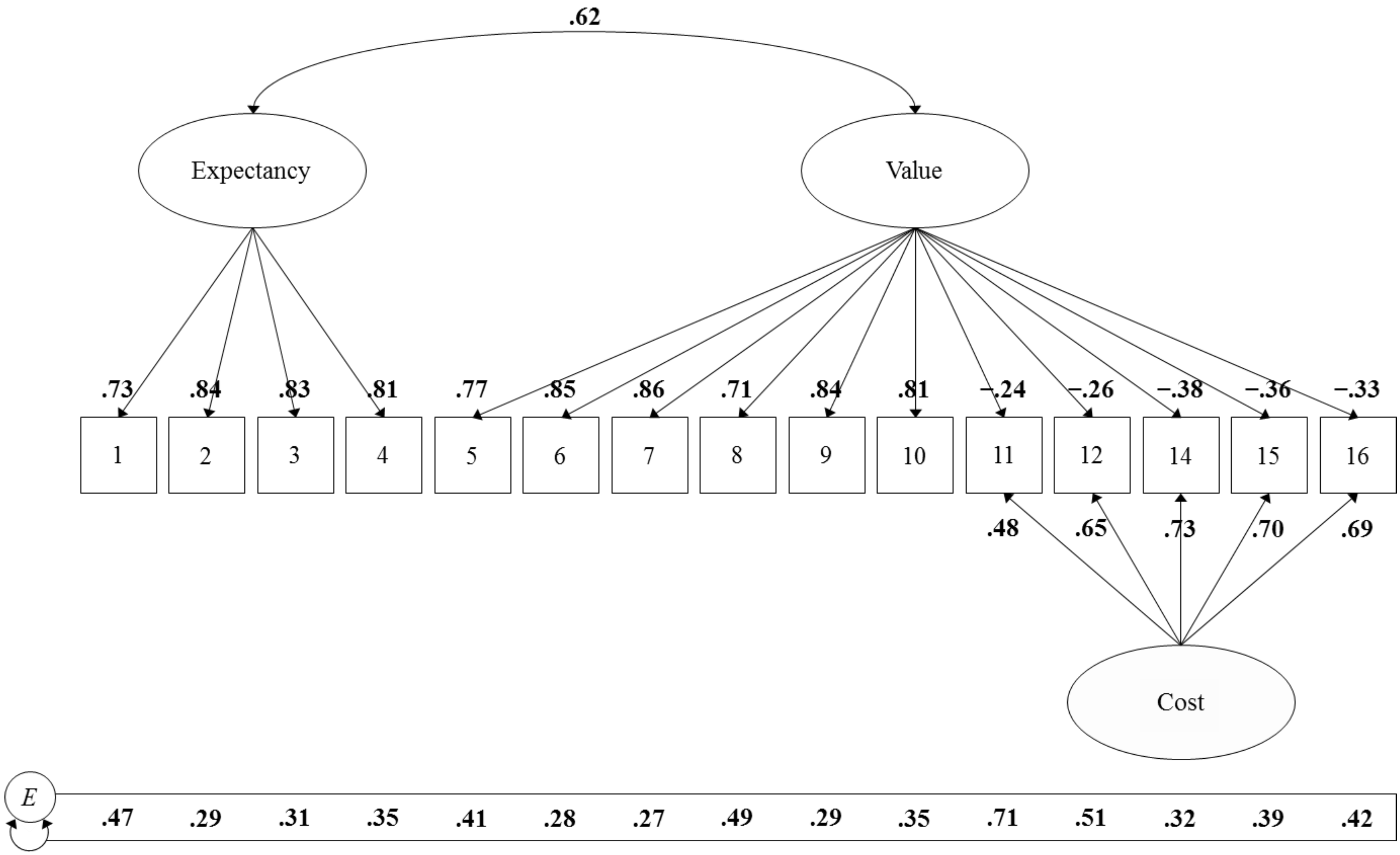


*Figure 5.* Standardized pattern coefficients and error variances in the 15-item bifactor model, Sample 1.

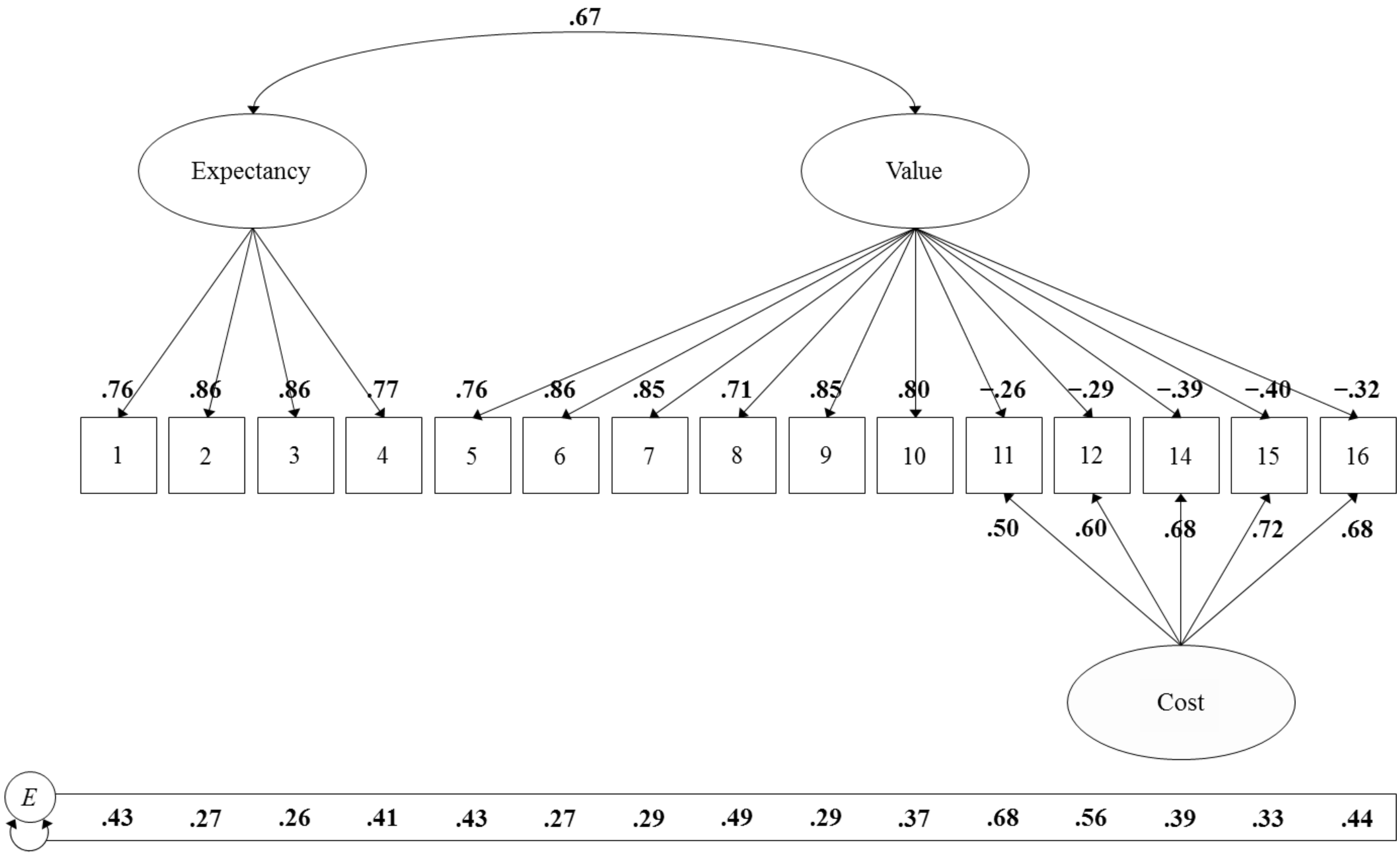


*Figure 6.* Standardized pattern coefficients and error variances in the 15-item bifactor model, Sample 2.

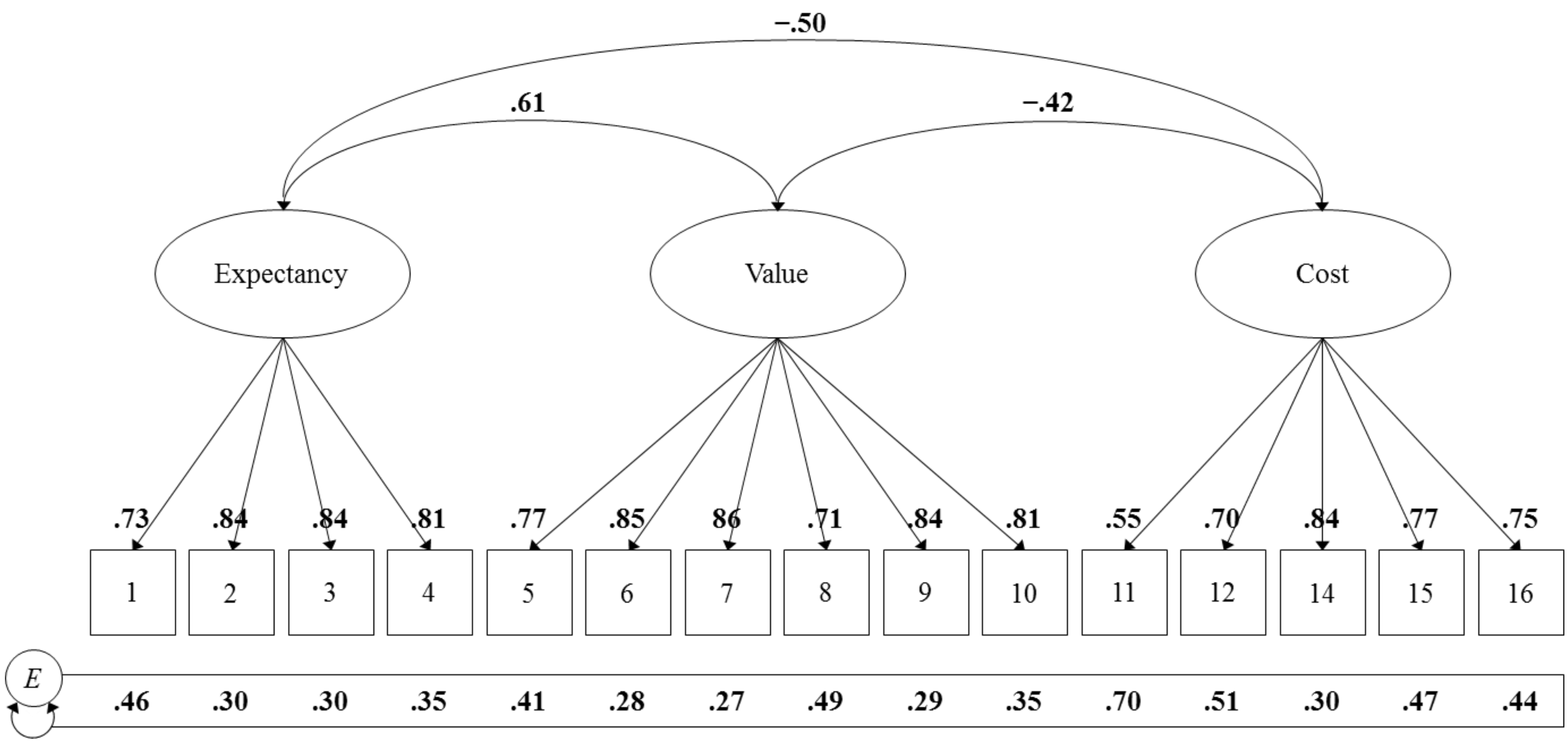


*Figure 7.* Standardized pattern coefficients and error variances in the 15-item three-factor model, Sample 1.

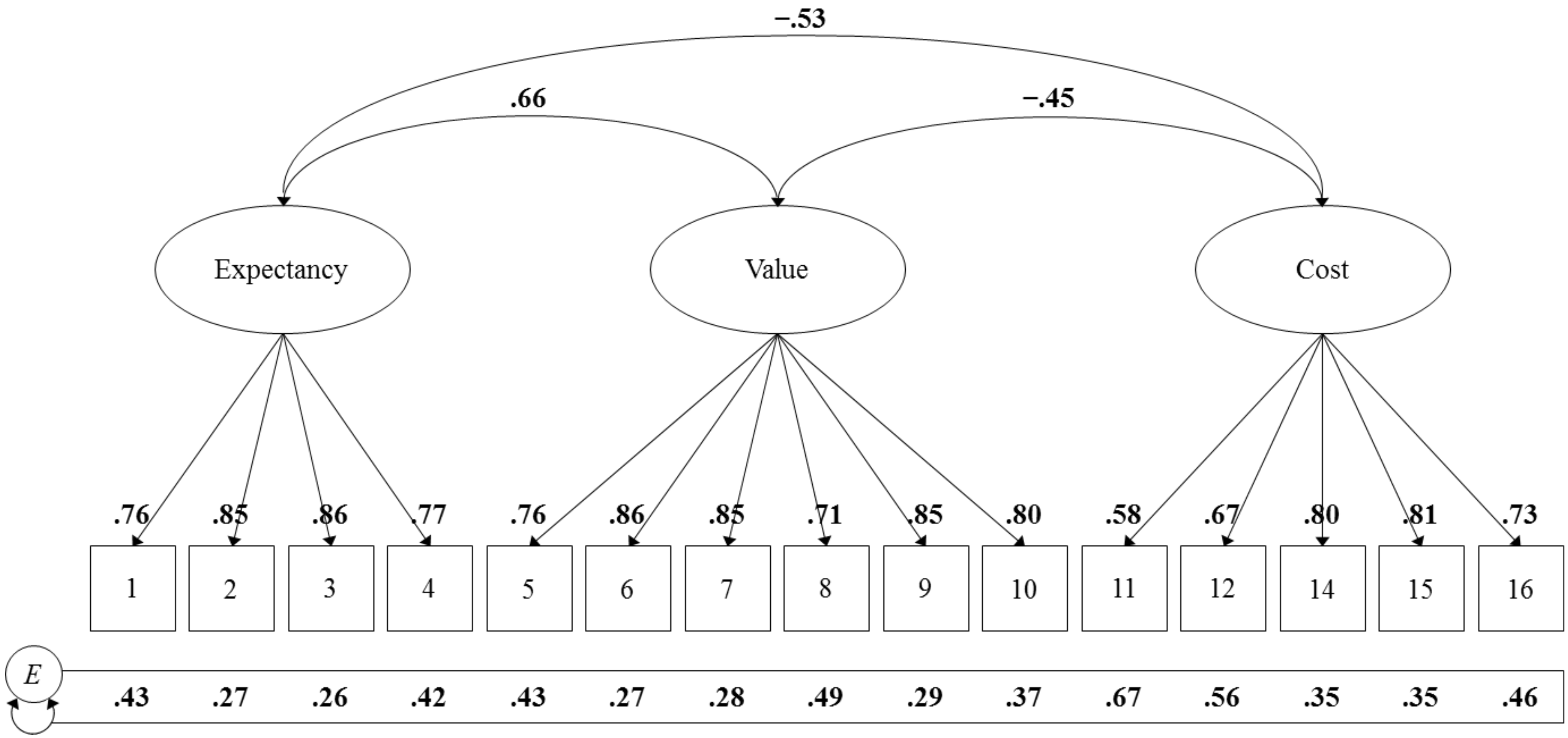


*Figure 8.* Standardized pattern coefficients and error variances in the 15-item three-factor model, Sample 2.

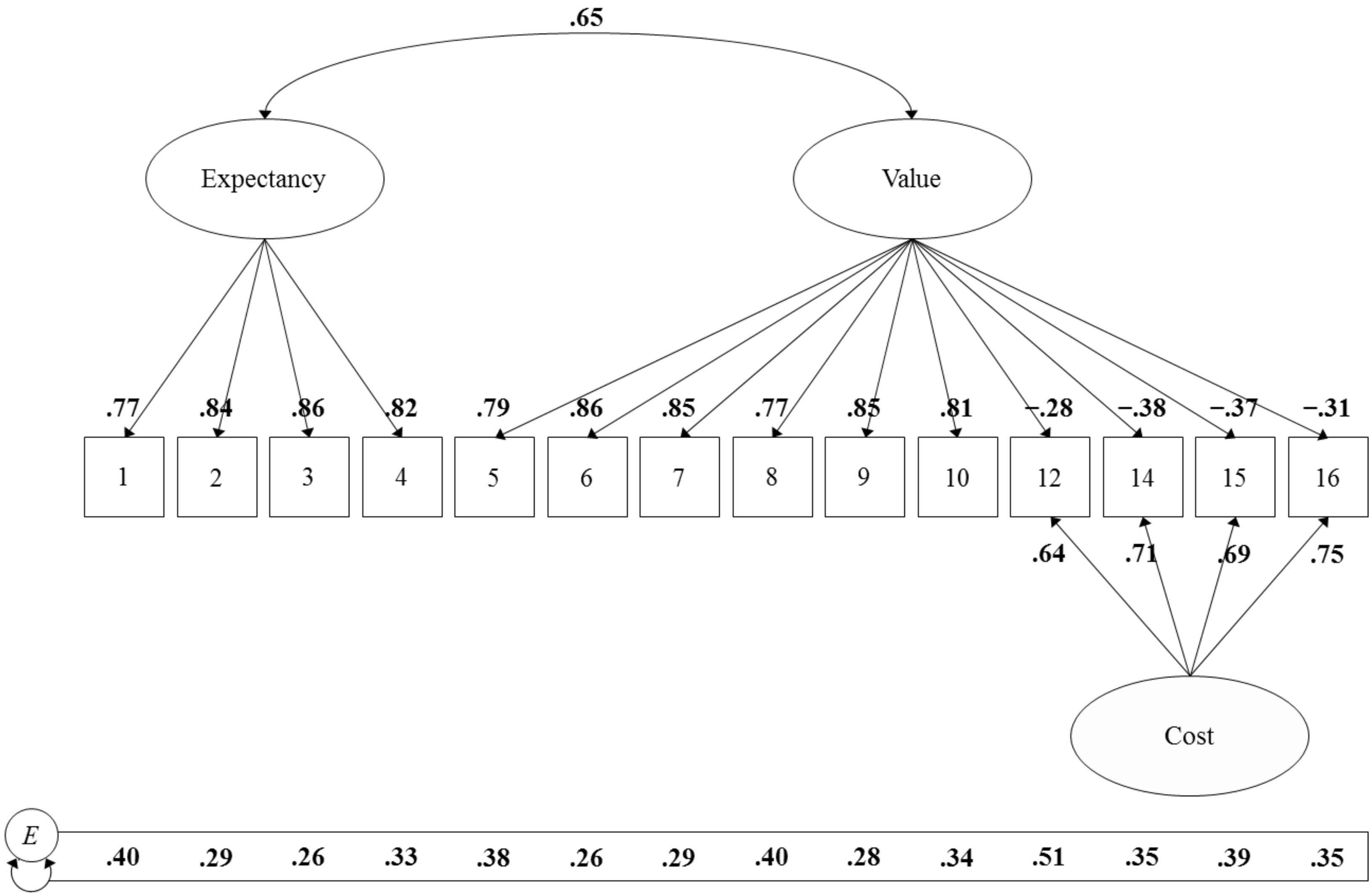


*Figure 9.* Standardized pattern coefficients and error variances in the 14-item bifactor model, Sample 3.

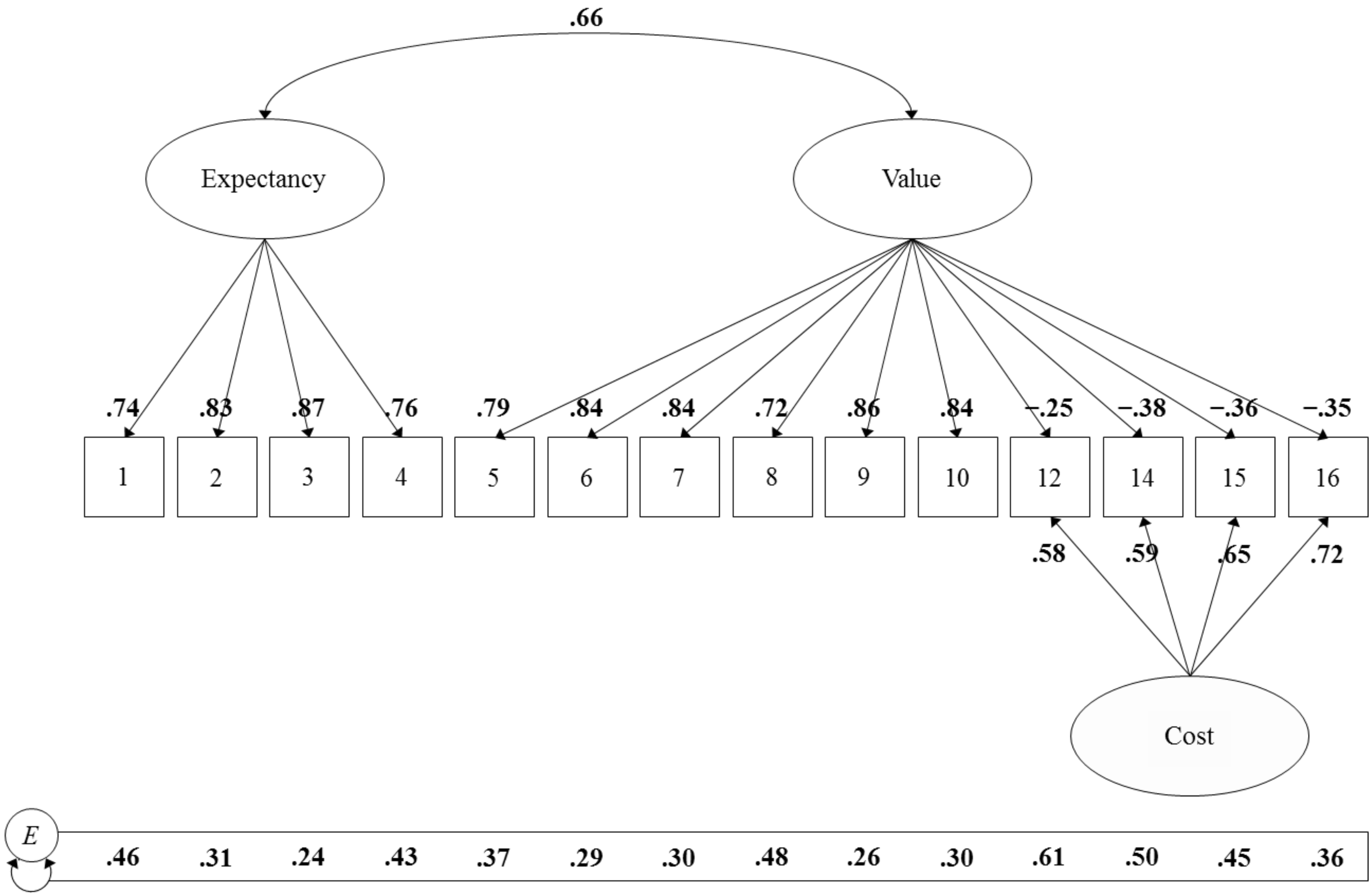


*Figure 10.* Standardized pattern coefficients and error variances in the 14-item bifactor model, Sample 4.

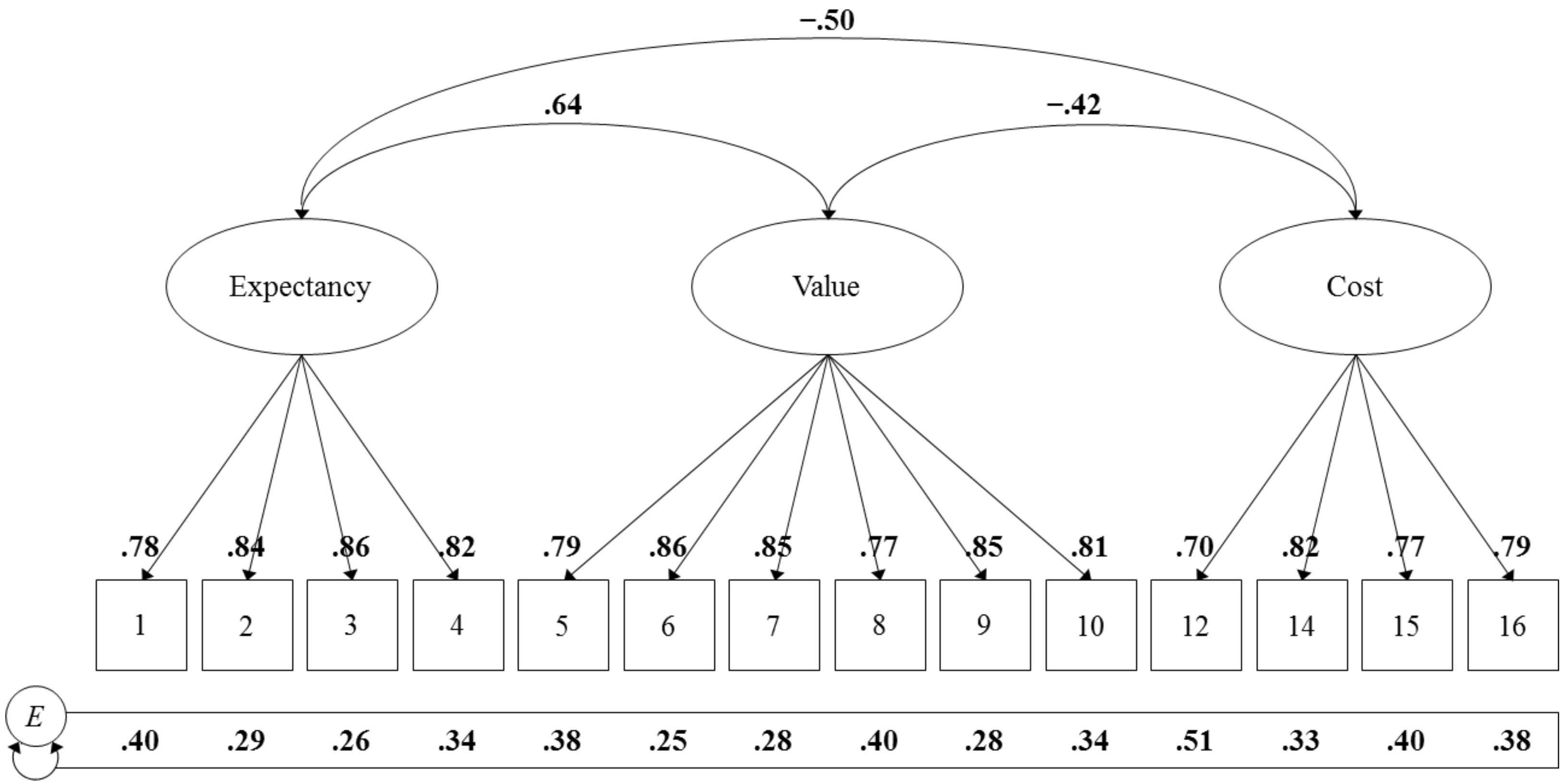


*Figure 11.* Standardized pattern coefficients and error variances in the 14-item three-factor model, Sample 3.

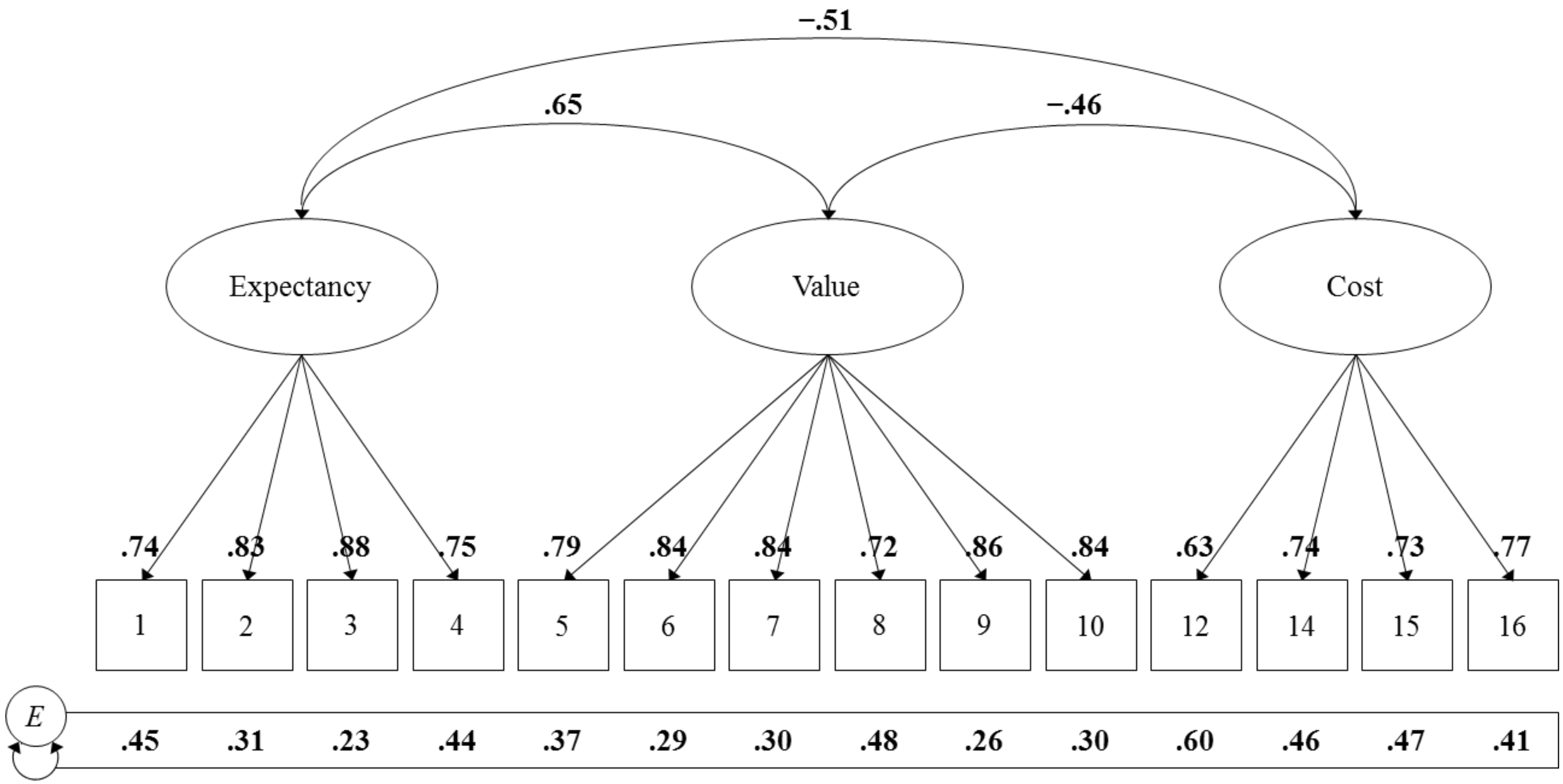


*Figure 12.* Standardized pattern coefficients and error variances in the 14-item three-factor model, Sample 4.

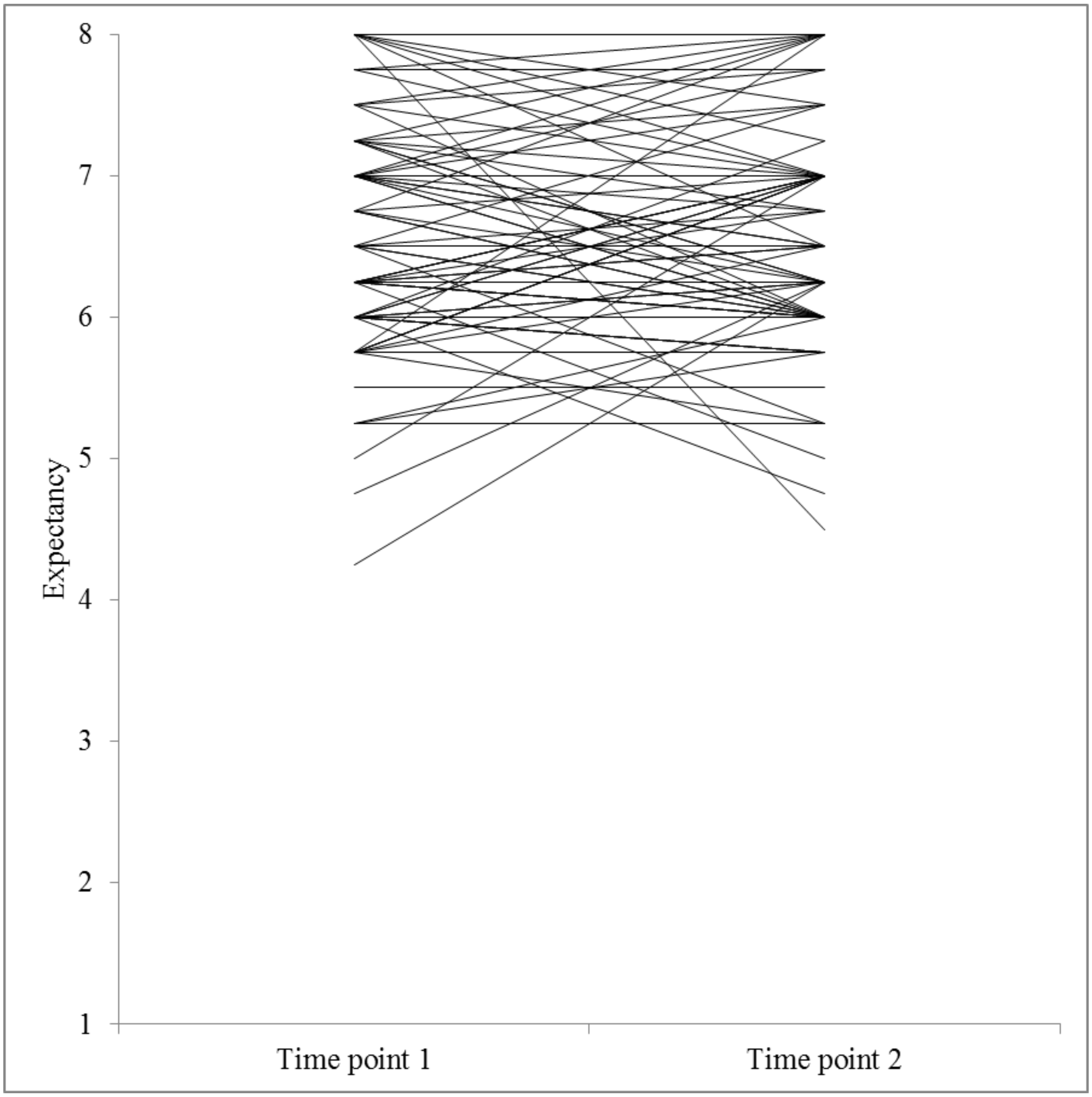


*Figure 13.* Individual change in Expectancy from time point 1 to time point 2 for a random sample of 91 (10%) students. Thirty-four students had decreases, 34 students had increases, and 23 students had no change in Expectancy between time points.

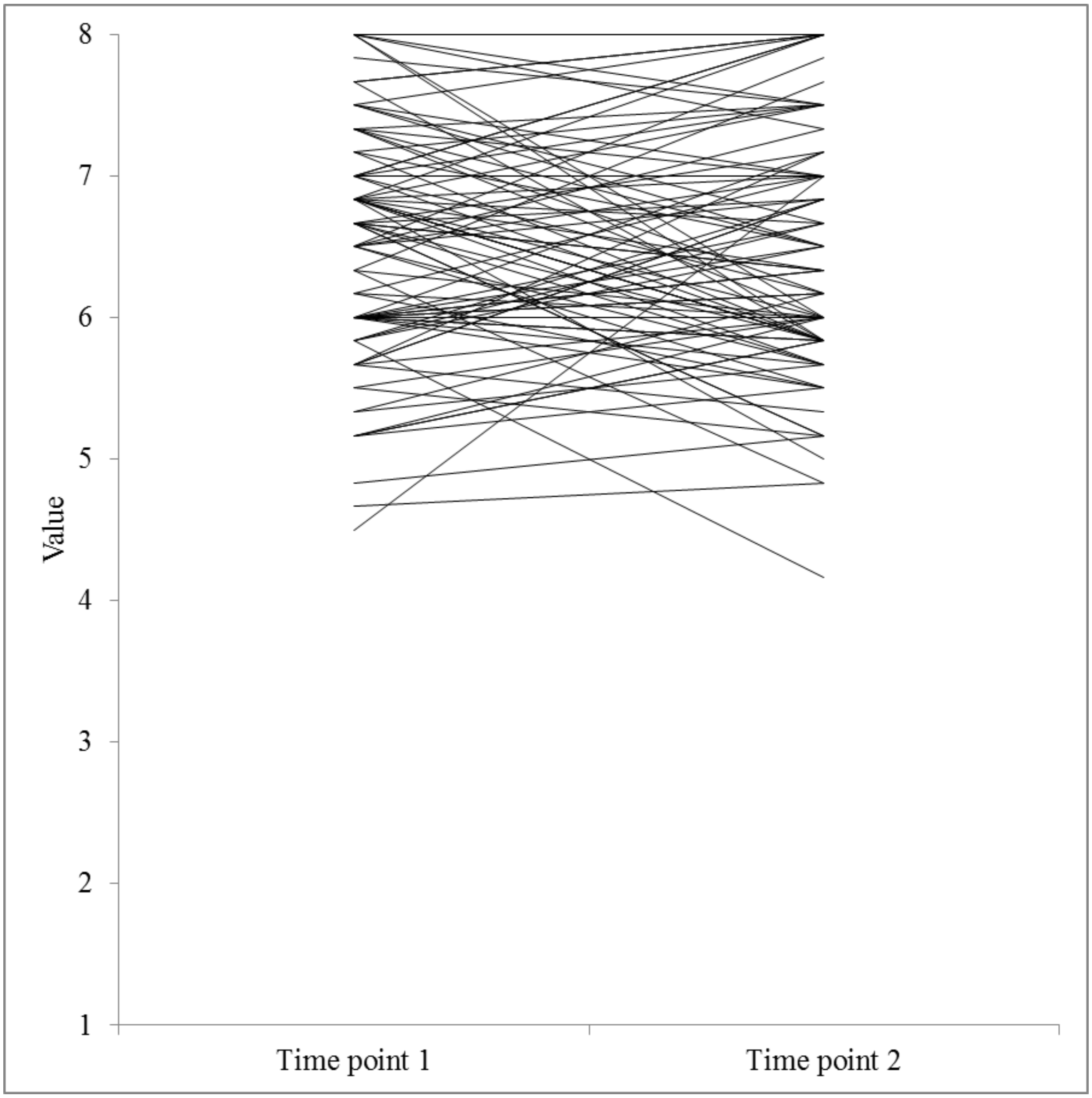


*Figure 14.* Individual change in Value from time point 1 to time point 2 for a random sample of 91 (10%) students. Forty-four students had decreases, 43 students had increases, and 4 students had no change in Value between time points.

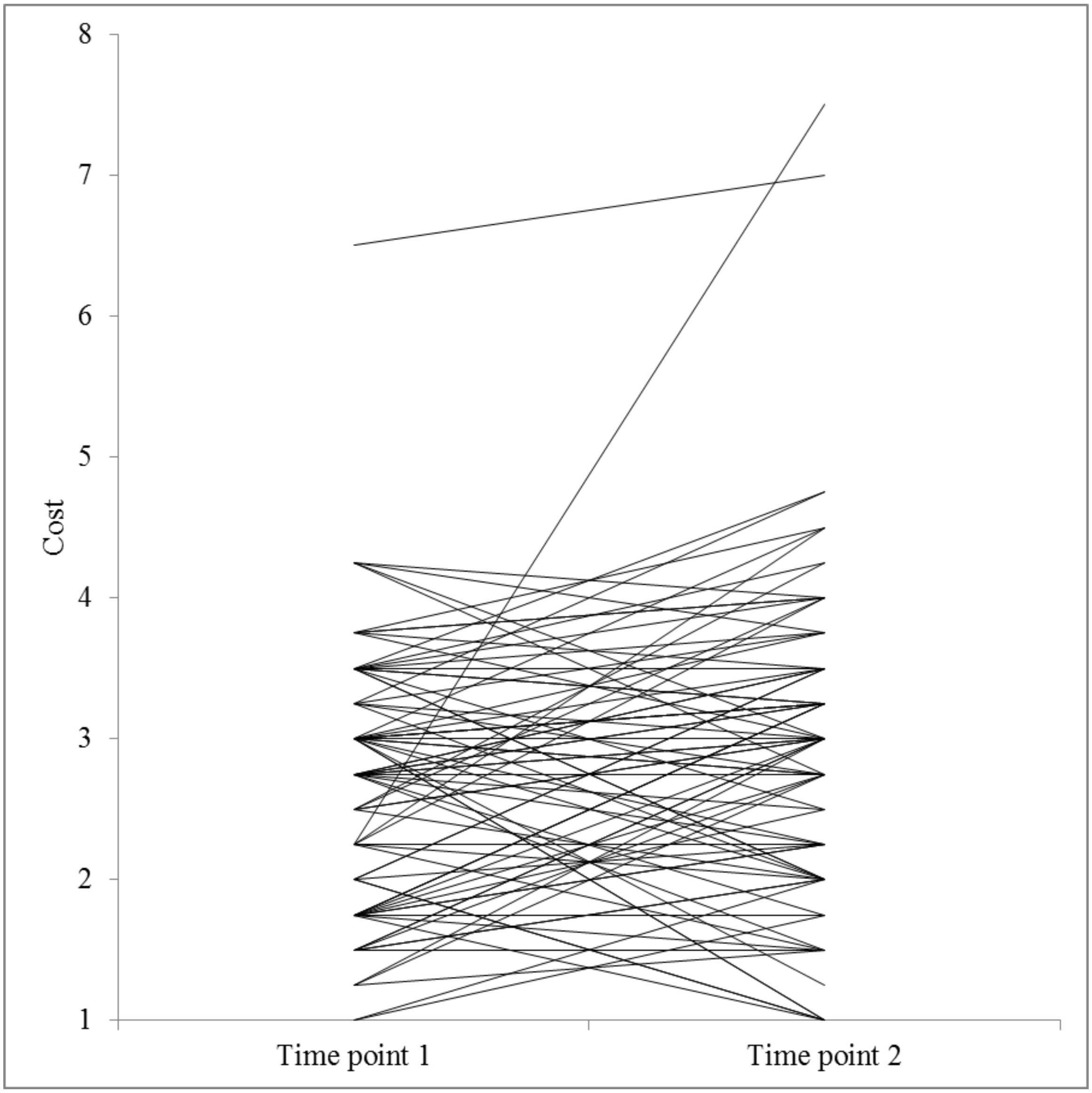


*Figure 15.* Individual change in Cost from time point 1 to time point 2 for a random sample of 91 (10%) students. Thirty-two students had decreases, 50 students had increases, and 9 students had no change in Cost between time points.

References

Adams, C. (2012, August 15). Student misconceptions about college life can undermine success. *Education Week.* Retrieved from http://blogs.edweek.org/edweek/college_bound/2012/08/student_misconceptions_about_college_life_add_to_challenge.html

Association of American Colleges and Universities. (n.d.). *What is a liberal education? And why is it important to my future?* Retrieved from http://leap.aacu.org/toolkit/wp-content/files_mf/leapstudentbroch_sample.pdf

Atkinson, J. W. (1957). Motivational determinants of risk taking behavior. *Psychological Review, 64*(6), 359-372.

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., … Zhang, J. (2012). *The Condition of Education 2012* (NCES 2012-045). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2012/2012045.pdf

Barron, K. E., & Hulleman, C. S. (2006). Is there a formula to help understand and improve student motivation? *Essays from E-xcellence in Teaching, Volume 8*. Retrieved from http://teachpsych.org/ebooks/eit2006/eit06-08.pdf

Barron, K. E., & Hulleman, C. S. (2010, May). *Introduction of the Motivation Research Institute for BSCS*. Invited seminar for the Biological Sciences Curriculum Study Research and Evaluation Team, Colorado Springs, CO.

Bashkov, B. & Finney, S.J. (in press). Applying longitudinal mean and covariance structures (LMACS) analysis to assess construct stability over two time points: An example using psychological entitlement. *Measurement and Evaluation in Counseling and Development.*

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17*(1), 10-17.

Bentler, P. M., & Wu, E. J. C. (2003). *EQS for Windows User's Guide*. Encino, CA: Multivariate Software, Inc.

Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *Journal of Educational Psychology, 93*(1), 23-34.

Bong, M. (2004). Academic motivation in self-efficacy, task value, achievement goal orientations, and attributional beliefs. *The Journal of Educational Research, 97*(6). 287-298.

Boning, K. (2007). Coherence in general education: A historical look. *The Journal of General Education, 56*(1), 1-16.

Bourke, B., Bray, N. J., & Horton, C. C. (2009). Approaches to the core curriculum: An exploratory analysis of top liberal arts and doctoral-granting institutions. *The Journal of General Education*, *58*(4), 219-240.

Boyer, E. L. (1987). *College: The undergraduate experience in America.* New York: Harper & Row.

Brint, S., Proctor, K., Murphy, S. P., Turk-Bicakci, L., & Hanneman, R. A. (2009). General education models: Continuity and change in the U.S. undergraduate curriculum, 1975-2000. *The Journal of Higher Education, 80*(6), 605-642.

Bui, K. V. T. (2002). First-generation college students at a four-year university: Background characteristics, reasons for pursuing higher education, and first-year experiences. *College Student Journal, 36*(1), 3-11.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456-466.

Conley, A. M. (2012). Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology*, *104*(1), 32-47.

Côté, J. E., & Allahar, A. L. (2011). *Lowering higher education: The rise of corporate universities and the fall of liberal education.* Toronto: University of Toronto Press.

Cuseo, J. (2005). “Decided,” “undecided,” and “in transition”: Implications for academic advisement, career counseling, & student retention. In R. Feldmen (Ed.), *Improving the first year of college: Research and practice* (pp. 27-48). Mahwah, NJ: Lawrence Erlbaum Associates.

DeCarlo, L. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*(3), 292-307.

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75-146). San Francisco: W. H. Freeman.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents’ achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215-225.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53,* 109-132.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age.

Finney, S. J., Pieper, S. L., & Barron, K. E., (2004). Examining the psychometric properties of the Achievement Goal Questionnaire in a general academic context. *Educational and Psychological Measurement, 64*(2), 365-382.

Fischer, K. (2013). A college degree sorts job applicants, but employers wish it meant more. *The Chronicle of Higher Education.* Retrieved from http://chronicle.com/article/A-College-Degree-Sorts-Job/137625/#id=overview

Flake, J. K. (2012). *Measuring cost: The forgotten component of expectancy value theory* (Unpublished master’s thesis). James Madison University, Harrisonburg, VA.

Flake, J. K., Barron, K. E., Hulleman, C. S., Lazowski, R. A., Grays, M. P., & Fessler, D. (2011, May). *Evaluating cost: The forgotten component of expectancy-value theory.* Poster session presented at the annual convention of the Association for Psychological Science, Washington, DC.

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling*, *13*(3), 378-402.

Gaff, J. G., & Davis, M. L. (1981). Student views of general education. *Liberal Education*, *67*(2), 112-123.

Gaff, J. G., & Wisescha, A. (1991). Assessing the reform of general education. *The Journal of General Education*, *40,* 51-68.

Getty, S., Hulleman, C. S., Barron, K. E., Stuhlsatz, A. M. & Marks, J. C. (2013, April). *Factors that affect learning in high school science: Measuring motivation, achievement, and interest in science*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Juan, Puerto Rico.

Gordon, V. N. (2007). *The undecided college student: An academic and career advising challenge* (3rd ed.). Springfield, IL: Charles C Thomas.

Gordon, V. N., & Steele, G. E. (2003). Undecided first-year students: A 25-year longitudinal study. *Journal of The First-Year Experience*, *15*(1), 19-38.

Grays, M. P., Hulleman, C. S., & Barron, K. E. (2012). *Students' motivation for the general education curriculum.* Manuscript in preparation.

Hart Research Associates (2010). *Raising the bar: Employers' views on college learning in the wake of the economic downturn.* Washington, DC: Hart Research Associates. Retrieved from http://www.aacu.org/leap/documents/2009_EmployerSurvey.pdf

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453.

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

Hulleman, C. S., Durik, A. M., Schweigert, S. A., & Harackiewicz, J.M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology, 100*(2), 398-416.

Johnston, J. S., Jr., Reardon, R. C., Kramer, G. L., Lenz, J. G., Maduros, A. S., & Sampson, J. P., Jr. (1991). The demand side of general education: Attending to student attitudes and understandings. *The Journal of General Education, 40*, 180-200.

Jöreskog, K. G., & Sörbom, D. G. (2006). LISREL 8.80 [Computer software]. Lincolnwood, IL: Scientific Software.

Kennett, D. J., Reed, M. J., & Lam, D. (2011). The importance of directly asking students their reasons for attending higher education. *Issues in Educational Research, 21*(1), 65-74.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher & C. Schatschneider (Eds.), *Applied quantitative analysis in the social sciences* (pp. 171-207). New York: Routledge.

Koljatic, M., & Kuh, G. D. (2001). A longitudinal assessment of college student engagement in good practices in undergraduate education. *Higher Education*, *42*(3), 351-371.

Kosovich, J. J. (2013). *Assessing intervention fidelity in a randomized field experiment: Illuminating the black box* (Unpublished master's thesis). James Madison University, Harrisonburg, VA.

Kosovich, J. J., Hulleman, C. S., & Barron, K. E. (2013, May). *Testing conflicting factor structures of expectancy-value*. Poster presented at the annual meeting of the Association for Psychological Science, Washington, DC.

Luttrell, V. R., Callen, B. W., Allen, C. S., Wood, M. D., Deeds, D. G., & Richard, D. C. S. (2010). The mathematics value inventory for general education students: Development and initial validation. *Educational and Psychological Measurement, 70*(1), 142-160.

MacCallum, R., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504.

Marsh, H. W., & Hau, K. T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, *32*(1), 151-170.

Miller, B. J., & Sundre, D. L. (2008). Achievement goal orientation toward general education versus overall coursework. *The Journal of General Education, 57*(3), 152-169.

Mulaik, S. A. (2009). *Linear causal modeling with structural equations.* Boca Raton, FL: Chapman & Hall/CRC.

Mulcahy, D. G. (2009). What should it mean to have a liberal education in the 21st century? *Curriculum Inquiry, 39*(3), 465-486.

Perk, S. (2005, October 13). On schooling's useless lessons. *The Daily Iowan.* Retrieved from http://dailyiowan.lib.uiowa.edu/DI/2005/di2005-10-13.pdf

Petrosko, J. M. (1992). *Measuring first-year college students on attitudes toward general education outcomes.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN.

Pew Research Center (2011). *Is college worth it? College presidents, public assess value, quality and mission of higher education.* Washington, DC: Pew Research Center. Retrieved from http://www.pewsocialtrends.org/files/2011/05/higher-ed-report.pdf

Phinney, J. S., Dennis, J., & Osorio, S. (2006). Reasons to attend college among ethnically diverse college students. *Cultural Diversity and Ethnic Minority Psychology, 12*(2), 347-366.

Pieper, S. (2003). Refining and extending the 2 x 2 achievement goal framework: Another look at work-avoidance (Doctoral dissertation, James Madison University, 2003). *Dissertation Abstracts International, 64,* 4436A.

Pracz, A. (2011, April 13). General education courses are a waste of time and money. *Northern Star.* Retrieved from http://northernstar.info/opinion/columnists/article_9c52826c-6639-11e0-8c12-0019bb30f31a.html

Pryor, J. H., Eagan, K., Palucki Blake, L., Hurtado, S., Berdan, J., & Case, M. H. (2012). *The American freshman: National norms fall 2012*. Los Angeles: Higher Education Research Institute, UCLA.

Reardon, R. C., Lenz, J. G., Sampson, J. P., Jr., Johnston, J. S., Jr., & Kramer, G. L. (1990). *The "demand side" of general education: A review of the literature* (Technical Report No. 11). Tallahassee, FL: Center for the Study of Technology in Counseling and Career Development, The Florida State University.

Sander, L. (2013, January 24). Freshman survey: This year, even more focused on jobs. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Freshman-Survey-This-Year/136787/

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent Variables Analysis* (pp. 399-419). Thousand Oaks, CA: Sage.

Selingo, J. (2013, February 21). Are career-oriented majors a waste of a 4-year higher education? *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/next/2013/02/21/are-career-oriented-majors-a-waste-of-a-4-year-higher-education/

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90.

Tabachnick, B. G., & Fidell, L. S. (2005). *Using multivariate statistics* (5th ed.). Boston: Allyn and Bacon.

Thompson, M. S., & Green, S. B. (2006). Evaluating between-group differences in latent variable means. In G. R. Hancock & R. O. Mueller (Eds.), *A Second Course in Structural Equation Modeling* (pp. 119-169). Greenwich, CT: Information Age.

Trautwein, U., Marsh, H. W., Nagengast, B., Ludtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy-value theory: A latent interaction modeling study. *Journal of Educational Psychology, 104*(3), 763-777.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-69.

Wade, S. (2013, January 14). General electives, generally a waste of time. *The Commonwealth Times.* Retrieved from http://www.commonwealthtimes.org/2013/01/14/general-electives-generally-a-waste-of-time/

Warner, D. B., & Koeppel, K. (2009). General education requirements: A comparative analysis. *The Journal of General Education, 58*(4), 241-258.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

White, E. R., & Cohen, J. (2004). Creating shared student responsibility for general education. *Peer Review, Fall 2004,* 8-10. Retrieved from http://www.aacu.org/peerreview/pr-fa04/pr-fa04feature3.pdf

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*(1), 49-78.

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*(1), 1-35.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*(3), 265-310.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68-81.

Yu, C., & Muthén, B. (2002, April). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Zavislak, A. (2012, October 9). Are general education requirements a waste? *The Cornellian.* Retrieved from http://www.thecornellian.com/2012/10/are-general-education-requirements-a-waste/

Zilberberg, A. (2013). *Students' attitudes toward institutional accountability testing in higher education: Implications for the validity of test scores.* (Unpublished doctoral dissertation). James Madison University, Harrisonburg, VA.