Measuring Motivation in Educational Settings: A Case for Pragmatic Measurement

Chapter · December 2017

3 authors:

Jeff John Kosovich
Center for Creative Leadership

Chris Hulleman
University of Virginia

Kenneth E. Barron
James Madison University

Measuring Motivation in Educational Settings: A Case for Pragmatic Measurement

Jeff J. Kosovich[1], Chris S. Hulleman[1], and Kenneth E. Barron[2]

[1]University of Virginia and [2]James Madison University

Measuring Motivation in Educational Settings: A Case for Pragmatic Measures

Critical to any empirical investigation is the quality of its measurement. Regardless of whether a researcher adopts a descriptive, correlational, experimental, quantitative, or qualitative approach to answer a given research question, valid measurement is essential for drawing meaningful conclusions from research. However, commonly accepted practices for ensuring measurement quality can be at odds with other situational constraints of a research study. Consider the three scenarios below that each highlight a tension between the technical standards of measurement and the practical constraints of the situation.

Scenario 1: Undergraduate students often enroll in five classes per semester to fulfill major, minor, general education, and degree requirements. *Question:* How does students' motivation differ across a variety of coursework and why might those differences exist? *Measurement Tension:* Can we collect data on students' motivation in a way that provides high quality data but does not frustrate students (i.e., does not require them to fill out the same lengthy survey for each class in which they are enrolled)?

Scenario 2: Research has demonstrated that student motivation declines from year to year starting in first grade and continuing through high school (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). *Question:* Are the patterns of motivational decline consistent throughout the school year or are there high points and low points? *Measurement Tension:* How can multiple types of motivation be measured frequently throughout the semester without disrupting class time and/or frustrating students?

Scenario 3: Motivation interventions can help improve student motivation and academic success (Lazowski & Hulleman, 2016). *Question:* If researchers were asked to develop a brief motivation intervention that could be delivered in a class in 15 minutes or less, could they

effectively change students' motivation? *Measurement Tension:* How can measures of motivation be included to evaluate the effectiveness of intervention in 15 minutes or less?

Each of these scenarios describes different real-world questions that measure users (i.e., researchers, practitioners, policymakers, measure developers) may have regarding student motivation. Each scenario also identifies tensions between the technical demands of high quality measurement and situational constraints that may make it difficult to answer the proposed questions. Although motivation plays a critical role in students' education, it is rarely assessed as often or as comprehensively as learning or achievement. Situational constraints in educational settings and applied research often limit how much, if any, information can be collected about students' motivation.

The purpose of our chapter is to describe the use of a flexible method of assessing measure quality in applied settings. We refer to this set of methods as *Pragmatic Measurement* (c.f. Kosovich, 2017, Paper 3). This approach seeks to minimize the resources required for data collection (e.g., class time, participant fatigue) while maximizing the quality of the data being collected (i.e., validity) (aka "practical measurement"; Yeager, Bryk, Muhich, Hausman, & Morales, 2013). We divide our chapter into three major parts. In Part 1, we focus on general approaches to measurement and provide an introduction on how pragmatic measurement can complement more common measurement practices. In Part 2, we elaborate on the scenarios above by considering three case studies where pragmatic measurement was used. In Part 3, we discuss the scholarly and practical implications of pragmatic measurement in the pursuit of educational motivation.

Part 1: Approaches to Measuring Motivation

What is Measurement and how is it Done?

The essence of measurement is to systematically and accurately capture a particular phenomenon (AERA, APA, & NCME, 2014). To do this, users need to ensure that responses to the measure correspond to the phenomenon under study—e.g., a measure designed to assess math self-efficacy should be distinct from items to assess content knowledge in math. (i.e., construct validity). Users also need to ensure that the measure elicits consistent responses—e.g., several different math self-efficacy questions should yield relatively similar responses from an individual (i.e., reliability). Ensuring that a measure is fulfilling its intended purpose is also known as *validation,* and validation for measures—whether they are self-report, behavioral, interviews, or otherwise—is the foundation of most research. However, the situational constraints of conducting measurement in applied settings (e.g., time constraints in classrooms) are often underappreciated in validation work, and can have far-reaching consequences for data collection, analyses, and conclusions drawn from the data. For example, in the first scenario described earlier (i.e., measuring motivation for five different classes), asking a student to fill out the same questionnaire five consecutive times would likely lead to frustration and/or disengagement. The frustration could be further exacerbated by lengthy questionnaires. Thus, when developing or adopting a particular measure, users must consider the quality of the measure as it fulfills technical standards (e.g., construct validity and reliability) and consider more pragmatic concerns (e.g., if completing the measures will frustrate participants).

When used for research purposes, measure development frequently occurs in several iterative steps (Schmeiser & Welch, 2006). First, items are identified that can be used to collect data (e.g., self-report questions, components of a checklist). The content of the items is reviewed by various experts for theoretical soundness, coherence, or other characteristics important for the scale (e.g., accurate translation from another language). Items are then piloted with potential

participants by directly completing the measure and/or discussing the measure with developers. For quantitative measures, various statistical procedures are conducted to assess whether the measure produces reliable scores (i.e., the items are consistent with each other or themselves over time), whether the items form distinct factors (i.e., items relate to each other more highly than they relate to items from other constructs), whether the items correlate or predict other theoretically important constructs in similar ways (i.e., correspond with similar constructs and diverge from less similar constructs), or whether the items predict outcomes (i.e., show that they meet a particular criterion). Qualitative measures require rigorous training of users in order to minimize bias and ensure proper measure deployment. Measures frequently undergo these and more complex processes multiple times during the course of development until a final useable measure is developed.

This common approach to measurement, which focuses on meeting the technical demands of providing high quality validity evidence, does not generally account for the practical constraints of a typical classroom. The resulting measures may be able to produce high quality information under ideal or controlled circumstances, but may breakdown once introduced to applied settings. The goal of this chapter is to demonstrate that useful information can be gleaned from measures that have balanced these technical and pragmatic considerations of conducting measurement in applied settings.

Pragmatic Measurement

Although the groundwork for pragmatic measurement exists in contemporary measurement theory, common measurement practices frequently adhere to outdated perspectives. One common example of this outdated knowledge is an exaggerated reliance on Cronbach's alpha coefficients (Sijtsma, 2009). Cronbach's alpha is one of the most widely cited and

reported reliability coefficients, yet there are volumes of research dedicated to debunking its usefulness. Even if a more appropriate coefficient is used, at least two items are required for their calculation (Traub & Rowley, 1991). However, the underlying goal of calculating reliability coefficients is not to achieve a specific number (as many users are inclined), but to demonstrate evidence of consistency (AERA et al., 2014). The over-reliance on alpha and other outdated perspectives can severely limit what validation methods are available to users interested in conducting measurement.

Pragmatic measurement is an approach to ensure that sound data on specific phenomenon can be collected under variable and constrained circumstances, especially when widely accepted validation methods may be unreasonably burdensome (Kosovich, 2017). We formally define pragmatic measurement as balancing psychometric concerns and situational constraints to produce maximally informative and minimally intrusive measures. For example in one study, faculty indicated that the researchers could spend no more than three minutes of class time per week conducting measurement (Yeager et al., 2013). Such constraints could be prohibitively restrictive if researchers are unwilling to compromise on either the number of constructs measured, the number of items dedicated to any given construct, or the mode of measurement. Although the measurement standards (AERA et al., 2014) acknowledge situational constraints, the role of those factors is frequently underappreciated in common practice.

Pragmatic measurement is an argument-based approach to validation (Kane, 1992, 2013) that draws from contemporary perspectives to pursue the underlying goal of measurement. The elegance of Kane's argument-based approach is that it requires measure users to identify the intended purposes of the measure (e.g., to compare group differences), and to identify what assumptions they are making about the measure (e.g., that the measure is sensitive to group

differences) (for examples, see Bell et al., 2012; Goldstein & Flake, 2016). The pragmatic measurement perspective adds an additional step of explicitly identifying potential constraints that could undermine the use of the measure, and incorporating those constraints in selecting or designing, and deploying the measure. We chose to build the pragmatic measurement perspective from Kane's framework because it is derived from a highly influential line of validation philosophy (Cronbach, Lee & Meehl, 1955; Messick, 1989) that guides contemporary measurement recommendations (AERA et al., 2014). Being a pragmatic perspective, we also adopted Kane's framework because of its flexibility as a general validation tool (i.e., argument-based validation is not limited to a single type of measure, construct, or domain).

For example, the authors' previous work (Kosovich, 2017; Kosovich, Hulleman, & Flake, 2017) identified prediction as an important use of a set of motivation measures. As evidence to using shorter measures rather than longer measures, the authors compared different versions of the scale (i.e., longer and shorter composites) to evaluate the amount of information lost. As would be expected, the shorter measures tended to explain less variance in outcomes. However, the decrease in variance explained was small in general, and proportionally smaller than the number of items eliminated from the longer scale. Based on those results, the authors concluded that while the brief measures did not technically perform as well as the longer measures, the overall benefits of shortening the scales outweighed the cost (by a factor of eight to one).

Rather than finding the scale with the highest reliability or the lowest standard error, the goal of our research on pragmatic measurement is to find the scale that performs well enough technically and meets the constraints of the situation. In Part 2, we present three scenarios where users might experience a tension between situational constraints and technical demands of validation. The scenarios were selected to be representative of common uses of measurement in

education research. Briefly introduced in the beginning of the chapter (i.e., student motivation across multiple classes, changes in motivation over time, and a time-restricted online motivation intervention), each scenario contains a measurement tension due to particular situational constraints that could be resolved by adopting the pragmatic measurement approach.

Part 2: Case Studies of Using Pragmatic Measures in Motivation Research

In the following sections, we consider three examples of pragmatic measurement in action. In the first scenario, the challenge from a pragmatic perspective was collecting data on multiple motivation constructs in five different domains in consecutive questionnaires. Not only would overly lengthy questionnaires likely lead to student disengagement, but questionnaire length was limited by the survey administrators. As a result, the researchers needed to produce brief measures that could capture a breadth of motivation constructs and a breadth of academic domains. In the second scenario, the challenge was collecting data several times in a single semester. Students tend to object to repeatedly responding to the same questionnaire, thus necessitating the usage of relatively brief measures. In the third scenario, the challenge was implementing an intervention and collecting data to assess the intervention in no more than 15 minutes. Again, brief measures were required and it was necessary to compile alternative validity evidence for the quality of the measures. These scenarios describe three common methodological approaches in motivational research (group differences, change over time, and interventions) and how the use of pragmatic measures of motivation can inform work on each.

Given the growing interest in motivation constructs, what kind of motivation should practitioners and policymakers assess? There are numerous motivational perspectives (Pintrich, 2003), which can certainly confuse those not familiar with the field. In an effort to integrate different facets of motivation (Atkinson, 1964; Lewin, Dembo, Festinger, & Sears, 1944;

Raynor, 1982; Vroom, 1964), an Expectancy-Value-Cost framework (Barron & Hulleman, 2015) offers a perspective on motivation that can be used to organize the array of motivation constructs (Eccles, 1983; Hulleman, Barron, Kosovich, & Lazowski, 2016). According to the Expectancy-Value-Cost framework, achievement behavior is most proximally determined by three major factors that can conceptually answer three questions (Barron & Hulleman, 2015). The first factor, e*xpectancy*, is an individual's perception that he or she can succeed at a task; it answers the question "Can I do this?" The second factor, v*alue* is an individual's perception that that engaging with a task is worthwhile; it answers the question, "Do I want to do this?" The third major factor, *cost*, is focused on individual's perception that there are psychological, physical, or other barriers preventing success; it answers the question "Is something preventing me from succeeding?"

Expectancies, values, and costs can be used as a first step when deciding what aspects of motivation are most salient in a particular context. For example, educators might predict that students would be most motivated in general education courses that are most similar to their college major (e.g., to a biology major, a science course would be more motivating than a classical literature course). It is likely that individuals would feel more confident (i.e., expectancy), find more meaning in (i.e., value), and be less anxious (i.e., cost) about courses that align with their interests or major. Without measuring motivation, it is difficult to tell if such a prediction is accurate. However, measuring motivation in detail is also probably unnecessary. The expectancy-value-cost framework historically employs brief measures to understand the role of motivation in education (Kosovich, 2017), and in each of the three examples below some version of a pragmatic measure of expectancy, value, and/or cost will be included. The examples

are meant to showcase the utility of pragmatic measures, not methods for designing pragmatic measures (for examples, see Kosovich, 2017; Yeager et al., 2013).

Scenario 1: Motivation for General Education

Post-secondary students are often required to take a variety of general education courses, many of which are not necessarily related to their major or their future plans in an obvious way. Because motivation is theorized to be (e.g., Eccles et al., 1983) and empirically shown to be (e.g., Jacobs et al., 2002) domain specific, students are more likely to be more motivated for some courses than for others. Presumably, students are more motivated for classes and activities related to their major. However, most research looking at student motivation across multiple domains does so for younger students and for more general topics (e.g., math, reading, sports). We were interested in answering two questions about college students' motivation: (a) How does students' motivation vary across their general education coursework? and (b) How does students' motivation for general education coursework compare to their major coursework? We explore the answer to these two questions using pragmatic measures of motivation in education (Barron, Grays, & Hulleman, 2014).

A representative sample of students from a large university (n = 808) were asked to respond to an 18-item pragmatic measure of motivation (see Appendix A). The measure asked about students' motivation in their major as well as in the five general education *areas* (i.e., academic domains). At the university under study, the five general education areas were (1) Critical Thinking and Communication Skills, (2) Arts and Humanities, (3) Math and Natural Science, (4) History and Social Science, and (5) Intrapersonal and Interpersonal Development. For each area and for their major, students were asked to rate their interest (e.g., "Area 1 courses interest me"), importance ("The topics and skills taught in Area 1 courses are important to me"),

and expectancy (“I can do well in my Area 1 courses”) for that area using a 1 (*Strongly Disagree)* to 6 (*Strongly Agree*) Likert scale.

Figure 1 presents the means and confidence intervals for students’ interest, importance, and expectancy across areas and for their major. Despite only using one item per construct and only measuring three constructs, a number of interesting patterns emerged from the data. First—as might be expected—students tend to report higher motivation for their majors than for the general education areas. Second, students report the lowest interest for Area 1 (Critical Thinking and Communication) and Area 3 (Math and Natural Science). Third, students report higher importance for Area 5 (Intrapersonal and Intrapersonal Development) than the other areas. Finally, Students report substantially lower expectancy for Area 3 than the other areas.

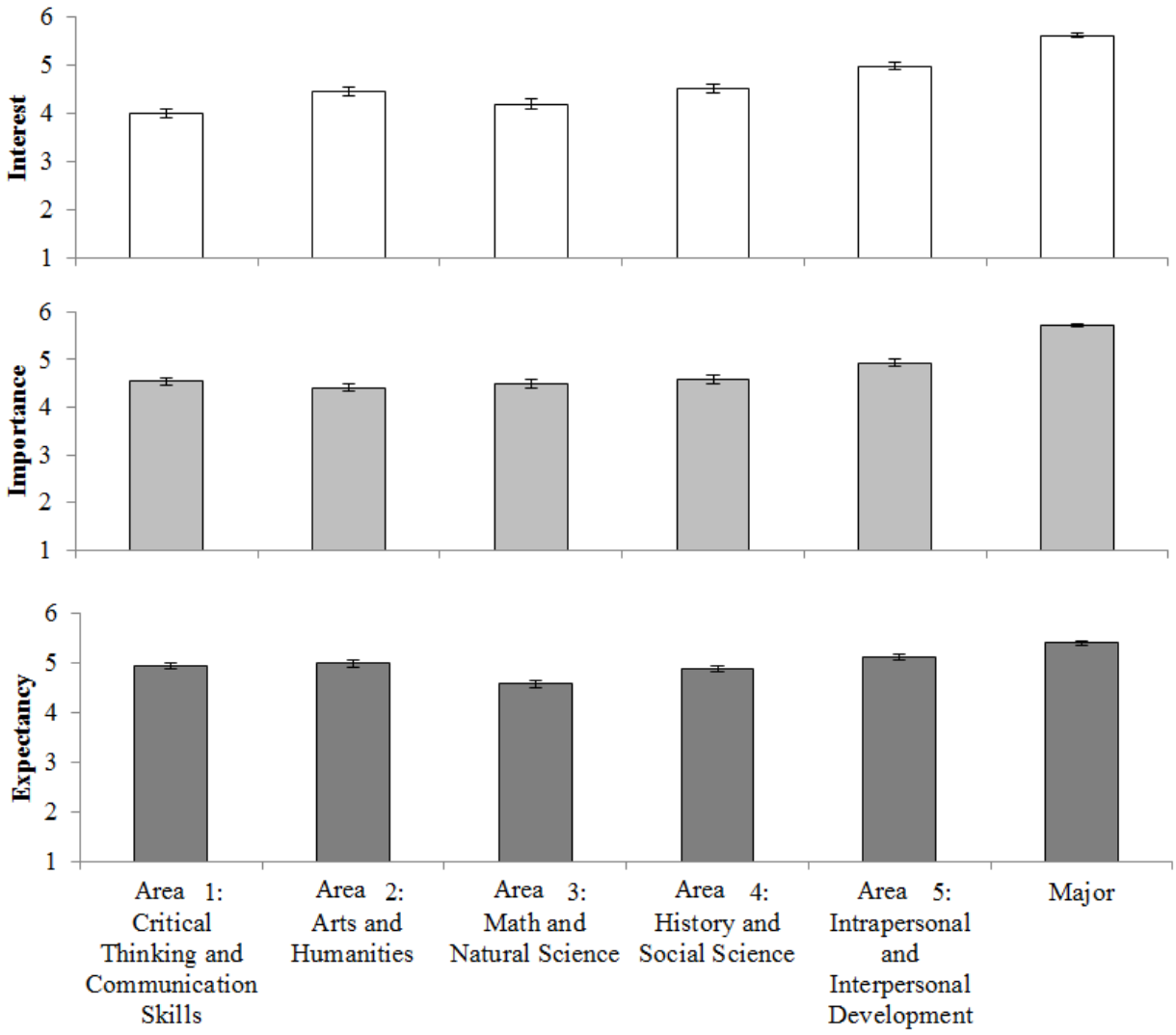

Figure 1. Students' self-reported interest (top), importance (middle), and expectancy (bottom) by general education area plus major. All measures used a six point Likert scale ranging from 1 (*Strongly Disagree*) to 6 (*Strongly Agree*).

Students were grouped into one of six curricular areas based on their declared major: (a) Arts & Letters, (b) Business, (c) Education, (d) Integrated Science and Technology, (e) Science and Math, and (f) Visual and Performing Art. A seventh group was comprised of Undeclared majors. The next step was to assess the extent to which students' motivation was higher for areas more similar to their major. For example, we expected that students with Science & Math majors

to give Area 3 (Math and Natural Science) the highest ratings on interest and importance because that area matches the coursework of their college and major. Students' ratings of different general education areas did differ by their major. For example, students with visual and performing arts majors reported higher motivation for Area 2 (Arts and Humanities) than any other major and also reported the lowest motivation for Area 3 (Math and Natural Science). Science and Math majors, along with Integrated Science and Technology majors, reported the exact opposite pattern, with the lowest motivation for Area 2 and the highest motivation for Area 3. Another interesting but also alarming result was that Education majors tended to have the lowest expectancy across areas, especially in Area 3 (Math and Natural Science).

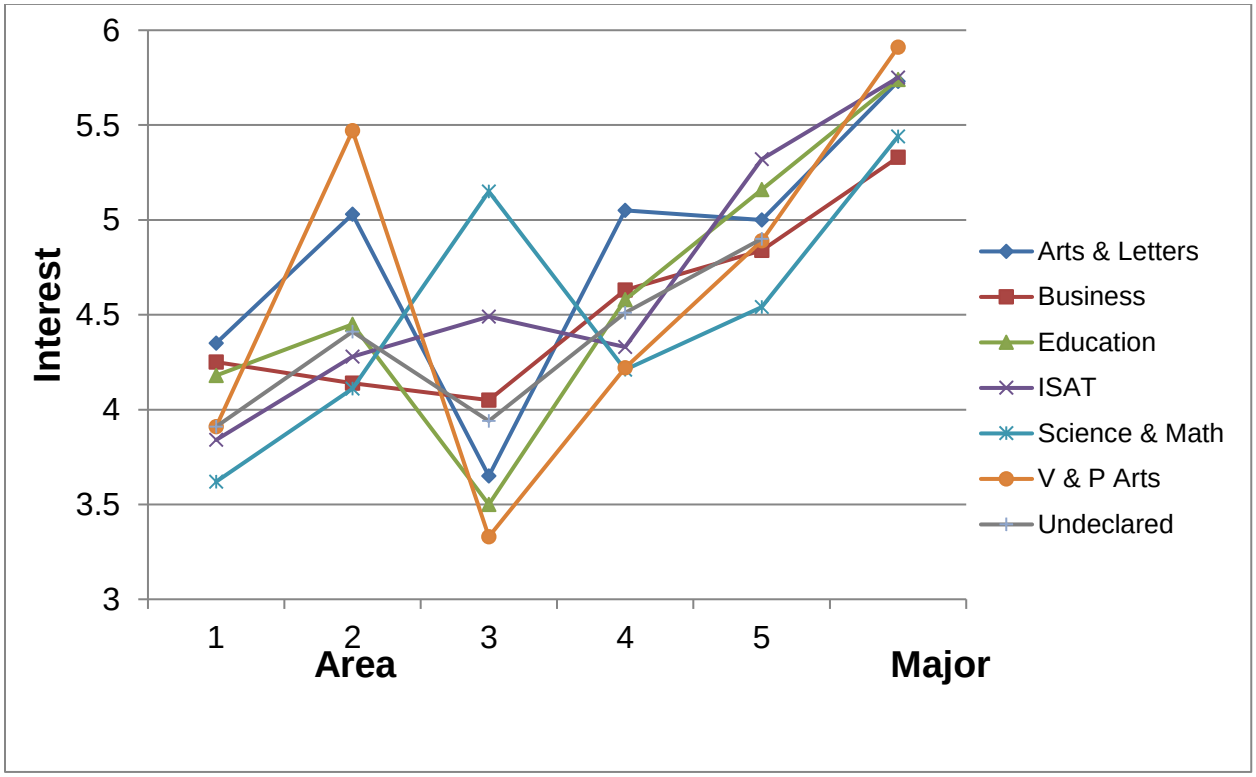

Figure 1

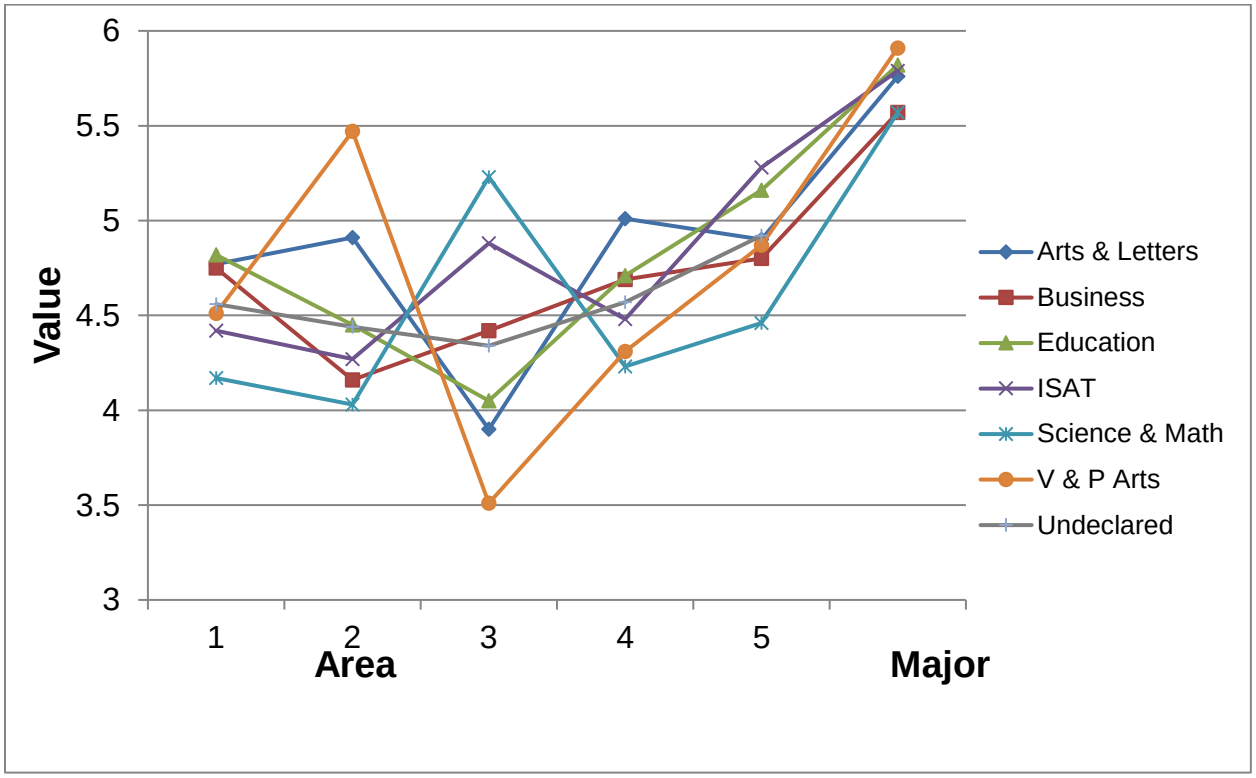
6
5.5
5
4.5
4
3.5
3
Value
1
2
3
4
5
Area
Major
Arts & Letters
Business
Education
ISAT
Science & Math
V & P Arts
Undeclared

Figure 2

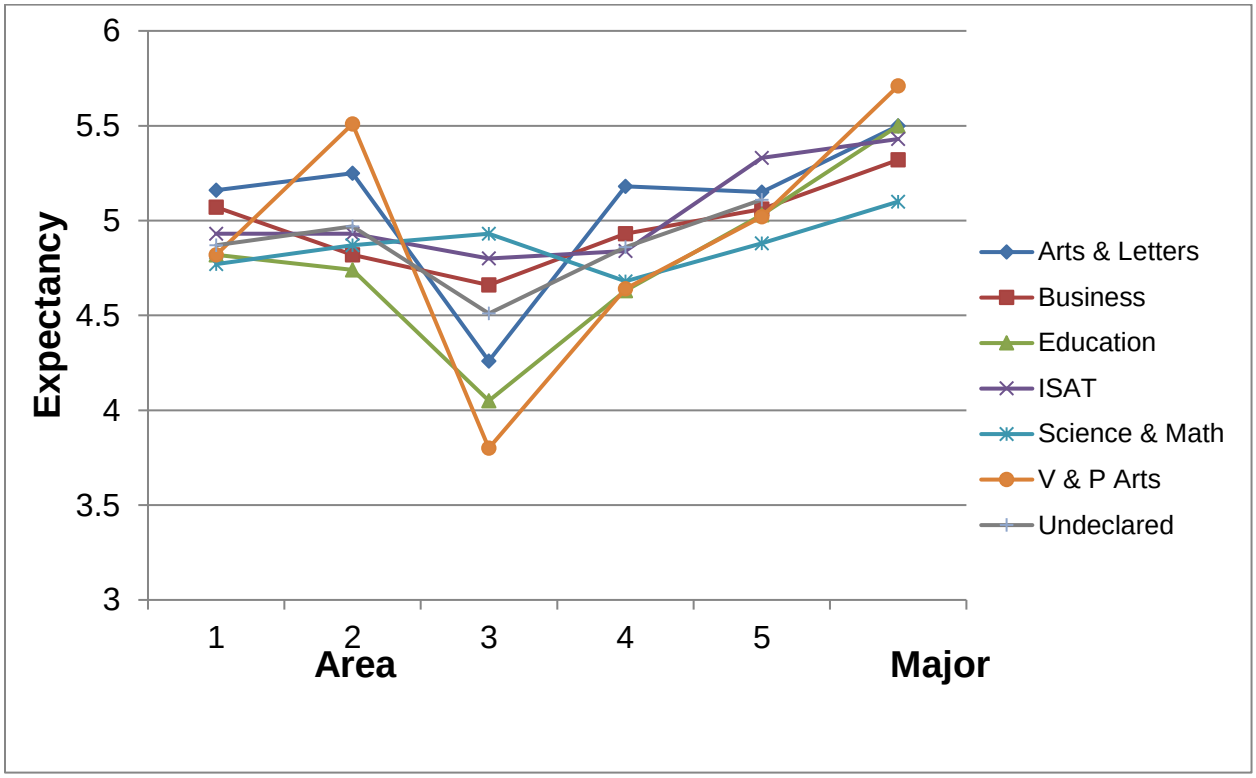

Figure 3

Scholarly signficance. The findings in this example offer a number of interesting and potentially important findings about motivation in higher education using single-item motivation measures. First, students' motivation for general education coursework varies by curricular area. Students from all majors indicated high motivation for Area 5 (Interpersonal and Intrapersonal Development) coursework, perhaps because students are able to see the relevance of studying intrapersonal and interpersonal phenomena beyond their major. Students tended to indicate the lowest interest in the foundational coursework comprising Area 1 (Critical Thinking and Communication), the lowest importance for coursework in the arts and humanities, and the lowest expectancy for math and natural science coursework. Second, students' motivation for general education was indeed lower than their motivation for their major coursework. Third, there was significant variability in attitudes toward Area 2 and Area 3, where higher interest,

importance, and expectancy were found among students whose majors were most closely aligned with the area coursework. Therefore, not only is it important to differentiate among components of the general education curriculum, but it is also important to recognize that attitudes toward each component are related to a student's academic major. These results highlight possible interventions to enhance students' motivation for their general education coursework. Area 2 showed the lowest importance overall, which means that many students—particularly those with a natural science bent—might benefit from an intervention that aims to increase value for the arts and humanities. Area 3 showed the lowest expectancy overall, indicating that many students might benefit from an expectancy-increasing intervention.

These relatively simple but striking results highlight an instance in which pragmatic measurement can contribute to discussions about education and much needed course reform efforts. These measures functioned as a brief assessment that could uncover patterns in this particular educational context. Although they do not provide explicit, diagnostic information, the data identify places that may be worthy of more lengthy assessments or other forms of inquiry (e.g., focus groups, observations).

Scenario 2: Short-Term Repeated Measures

Our next example demonstrates that pragmatic measures also can provide similar benefits when considering motivation over several points of time. Motivational theory suggests that motivation is dynamic and is a culmination of prior experiences and attitudes combined with present contexts (Eccles et al., 1983). Students who demonstrate low achievement are also likely to diverge motivationally from students who demonstrate high achievement. As a result, it is important to monitor motivation over time—how students feel about a class is likely to be different at the end of a semester than the beginning. For example, a common finding in the

motivational literature is that motivation declines over time. The pattern has been documented repeatedly, particularly in research from the expectancy-value framework that spans first grade through the end of high school (Jacobs et al., 2002).The implications of decreasing motivation for a particular domain means the individuals are less likely to pursue the domain in the future, are likely to perform worse in the domain, and are likely to have continued decline in their motivation. However, the majority of research examining motivational change occurs at a yearly-time scale and ignores shorter time intervals. A concern with more frequent measurement is that participants will become fatigued or disengage with the general research process.

Our second example addresses this gap in the literature by examining how students' expectancy and value change over the course of a semester, and whether that change is consistent across individuals or not (for the full study results, see Kosovich, Flake, & Hulleman, 2017). Students' (n = 389) motivation was measured three times over the course of a semester, during week 3 (Time 1), week 9 (Time 2), and week 14 (Time 3). We used eight items adapted from Eccles and colleagues (Eccles et al., 1983) to capture students' expectancy (5 items, e.g., "I expect to do well in this class) and value (3 items; e.g., "What I am learning in this class is relevant to my life."). Analyses were conducted using a parallel process model (Byrne, 2012) which examines the patterns of two or more constructs over time. The model provides *growth parameters* which include the level of each construct at the first time point (i.e., an intercept) and the rate of change in the construct for one time interval (i.e., a slope).

First, we found that the general declining trajectory in prior research was replicated in the current study. On average, expectancy and value both declined over the course of the semester. Importantly, the slopes were highly correlated $r = .83$, $p < .01$, suggesting that expectancy and value change in tandem. However, the variability in expectancy slopes was statistically

significant, while the variability in value slopes was not. In other words, not all individuals declined in their expectancies, some increased, some decreased, and some remained stable. Individuals were more homogenous in their value trajectory, but the change that was present was highly related to expectancy. Second, growth parameters were uncorrelated with final exam scores. However, value intercepts and expectancy slopes both predicted end-of-semester self-reported continuing interest. Thus, individuals with higher initial value showed greater interest at the end of the semester. Finally, concurrent exam scores were able to explain about 52% of the variation in expectancy slopes. Specifically, students who performed better on exams also reported higher expectancies.

Scholarly Significance

Although we know that motivation changes over the long-term, there is little research illuminating the short-term dynamics of motivation change (see Holstermann, Ainley, Grube, Roick, & Bögeholz, 2012 for an example). Even when researchers repeatedly measure motivation over extremely short periods of time, they typically do not analyze change or variation among the individual time points (e.g., Hektner & Csikszentmihalyi, 1996; Rotgans & Schmidt, 2011). The results of this example show that although expectancy and value indeed decline over the semester, the nature of that decline is complex. Variability in expectancy slopes was related to concurrent exam performance supporting the link between perceptions of ability and performance (e.g., Bandura & Schunk, 1981). Indeed, many students' expectancy increased despite an average decline. The lack of variability in value slopes was also interesting, suggesting either that value is less susceptible to feedback about value or feedback about value is simply not communicated to students. We know from our previous research that effective value feedback can lead to improvements in expectancy, value, interest, and performance later in the

semester (Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009; Hulleman, Kosovich, Barron, & Daniel, 2016). It is worth noting that expectancy and value measures can be considered pragmatic measures as they tend to be relatively brief. In the current example, eight items were used to measure expectancy and value, but fewer items have been used for the same types of analyses in other studies (e.g., Jacobs et al., 2002; Kosovich, 2017; Kosovich, Hulleman, Barron, & Getty, 2015; Kosovich, Hulleman, et al., 2017)

Perhaps most important are the effects of growth parameters on continuing interest. At a generic level, the results support hypothesized mechanisms for interest development and maintenance (Hidi & Renninger, 2006; Renninger & Hidi, 2016). First, Hidi and Renninger hypothesize that interest is more likely to develop and be sustained when individuals perceive meaningfulness in the content. Congruent with that hypothesis, we found that individuals who see greater usefulness in the topic at the beginning of the semester are more likely to maintain or increase interest. Similarly, interest maintenance and development requires support from external sources, such as competence feedback. Congruent with that hypothesis, we found that consistent change in expectancy predicted later interest, controlling for prior interest.

Scenario 3: Online Motivation Interventions

The first two examples demonstrated instances where pragmatic measurement could be used to monitor natural patterns in a situation and provide useful information to relevant stakeholders (e.g., researchers, administrators). The third example takes another step, leveraging pragmatic measurement as a way to include assessments despite significant situational constraints. Researchers, practitioners, and policymakers are continually introducing new programs, initiatives, and curriculum (e.g., educational interventions) in hopes that they will improve student learning outcomes. In recent years, researchers have demonstrated that utility

value interventions can increase students' interest in STEM, academic performance, and subsequent course taking (see Lazowski & Hulleman, 2015; Rosenzweig & Wigfield, 2016; Tibbetts, Harackiewicz, Priniski, & Canning, 2016, for reviews of this work). An important but difficult aspect of introducing and maintaining such interventions is assessing whether or not they bring about the expected changes in new contexts. Adopting interventions can be disruptive for classroom routines and can be time consuming. Including comprehensive assessments in addition to interventions may only add to the disruption and frustration. Once again, pragmatic measurement can provide a method for including useful measures without undermining the goal of the research or data collection. Utility value interventions tend to be relatively brief activities (e.g., 30 minutes) that ask students to write a short essay reflecting on how class material might be relevant or useful to their lives (Harackiewicz et al., 2014; Hulleman, 2007; Hulleman & Harackiewicz, 2009; Hulleman, Kosovich, et al., 2016; Hulleman, Schrager, Bodmann, & Harackiewicz, 2010). Recent versions have been adapted to be delivered online (Hulleman, Kosovich, et al., 2016). In this third scenario, the researchers were given 15 minutes to complete the intervention and all relevant assessments.

This example includes a sample of algebra and geometry students (n = 797) from an accredited online high-school in the southeastern United States. The classes were self-paced and consisted of approximately 10 topic modules for students to complete. Students received an e-mail from their course instructors, asking them to volunteer their time by completing an online activity about how math was related to their lives. The study procedure was as followed: 1) consent form, 2) a five-item baseline motivation questionnaire (self-efficacy and interest), 3) a 10-15 minute intervention, and 4) an eight-item follow-up questionnaire (self-efficacy, interest, utility value, and cost). We assessed participants' perceptions of utility value using one item,

interest was measured with three items (e.g., I am interested in taking more math classes in the future"; α = .89), self-efficacy was measured with two items (e.g., "*I am confident that I can figure out even the hardest concepts in algebra/geometry class*"; α = .90), and perceived cost was measured with two items (e.g., "*I'm unable to put in the time needed to do well in algebra/geometry*"; α = .87). Participants responded using a 5 point Likert-type scale from 1 (Not at All) to 5 (Extremely). Students who opted into the study were randomly assigned to a utility value intervention or a control activity (for full details see Rosenzweig et al., under review). The primary dependent measure was students' perceived utility value for the course. We also analyzed students' scores on the other motivational constructs to determine whether the interventions improved motivation broadly instead of specifically improving utility value**.**

To model the effects of the intervention conditions on the different outcomes, we ran a series of multiple hierarchical linear regressions. We regressed each outcome on a set of dummy-codes, which represented each condition relative to the control groups. We included students' course type, pre-test competence beliefs, and pre-test interest as covariates in the model. The results of the study showed that all three groups of students who received a utility value intervention reported higher utility value than the control groups (see Figure 4). Consistent with work by Gaspard and colleagues (2015), the condition in which individuals were asked to evaluate quotes rather than write an essay reported higher motivation after the intervention. There were no significant main effects of condition on post-test interest, cost, or self-efficacy, compared to either of the control conditions.

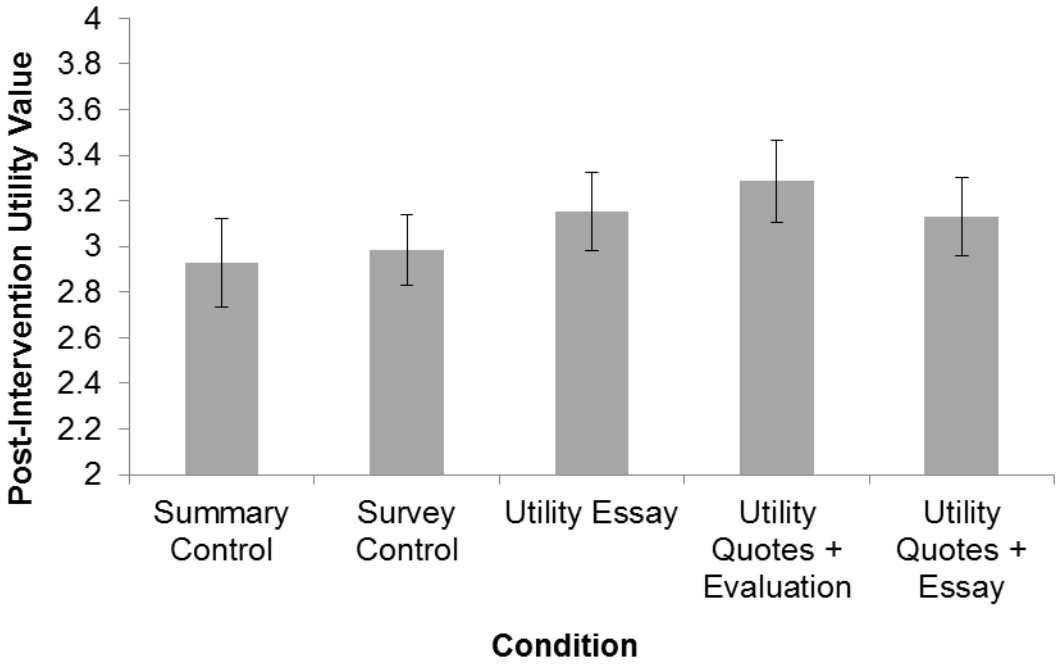

Figure 4.

Scholarly significance. In this research we explored whether we could raise students' perceived utility value for math in online math courses through utility value interventions. To our knowledge, this is one of the first studies to demonstrate that utility value interventions can increase high school students' perceived utility value for learning math in online courses. Relative to the control conditions, the utility intervention that provided examples and non-essay tasks (i.e., the Utility Quotes + Evaluation condition) increased students' utility value the most. Our findings support prior research suggesting that it is possible to raise students' perceptions of utility value by asking them to make specific and personal connections to their course material (e.g., Hulleman & Harackiewicz, 2009; Hulleman et al., 2010, 2016). We extend prior work to show that these interventions raise utility value for students in online courses, a growing segment of K-12 education that has been under-studied in intervention research.

Furthermore, this research explored the boundary for how brief can a utility value intervention be and still produce positive effects for students. In this case, completing the

intervention materials themselves took 5-10 minutes, which is one-third of the length of time of most other interventions, and less than one-tenth of the time of the Gaspard et al. (2015) intervention. This reinforces the notion that when interventions target specific psychological processes, that they can have powerful effects despite minimal dosage (Yeager & Walton, 2011).

Part 3: Scholarly and Practical Implications

The scenarios discussed in Part 2 highlight cases in which pragmatic measurement can help answer motivation questions. The perspective allows users to identify what information would support the use of the measure, rather than simply recommending a general checklist of requirements that may or may not be important. It also allows for more rapid measure development and use of measures because it prioritizes the most important evidence rather than a generalized checklist of requirements. Additionally, the pragmatic measurement perspective also gives individuals the opportunity to collect some data (rather than none) if they do not have the resources to implement or develop a full-length scale. Importantly, with this perspective we emphasize the usability of data that are *good* enough, recognizing that measures and data are rarely ever ideal for the applied researcher, practitioner, or policymaker. Finally, the perspetive can reduce the burden of frequent responding on participants by minimizing the demands of completing a measure.

Benefits and Applications of Pragmatic Measurement.

The first scenario illuminated patterns of motivation across five general academic areas, students major, and the correspondence between the two. Although the analyses were relatively simple, they were able to provide a substantial amount of information. This is one core use of pragmatic measures—to collect a breadth of information from a large group of individuals to inform practitioners. The results of the study in the first example could form a foundation for

researchers or practitioners to launch more targeted inquiries. For example, why are education majors consistently reporting lower expectancies than other majors across academic areas?

The second scenario represented a more nuanced look at motivation over time. The results of the study not only further theoretical knowledge, but bring to light important points for practitioners. For example, the uniformity of the decline in utility value suggests that students may share similar experiences in a psychology class that does not foster value for the class material. Tracking fluctuations would be one obvious use of pragmatic measures. The research design in the second scenario still only represented motivation at the beginning, middle, and end of the semester. A major question is how motivation fluctuates at a more micro level, such as from week to week. Experience sampling methods can be used to examine motivation change in more narrow time intervals (e.g., Hektner & Csikszentmihalyi, 1996), but that method is intense and can be demanding for participants. A balance of the two might be a weekly three item survey at the beginning or end of class. Such an exercise would be feasible with much larger sample sizes and could potentially be implemented at a much larger scale (e.g., a college or school district).

In the third scenario, we were able to assess motivation before and after an intervention within a 15-minute timeframe. The brief measures allowed us to quickly assess several potential motivation factors that may be important to online high school math students. The results of the study demonstrated that even a 15-minute exercise may be sufficient for altering student motivation. From a pragmatic stand point, these results could lead to negotiating for a longer intervention time slot, to gain more buy-in from skeptical instructors or administrators, or to shorten other lengthier versions.

Beyond these three applications, pragmatic measures should have additional utility for other purposes (see Table 1). One would be routinely using pragmatic motivation assessment in teaching practice. Our research suggests that at least some measures of motivation do not require advanced statistical modeling to lead to accurate conclusions (Kosovich et al., 2015). This could allow teachers or administrators to monitor motivation throughout the school year and develop curricular changes that might support motivation during common declines. Similarly, it could provide another source of information to identify students who are struggling with material and/or experiencing other issues. For example, students who report particularly high cost (e.g., feeling overwhelmed by their coursework) may benefit from time-management instruction or tutoring. If students have an anonymous outlet to inform instructors that they are not confident with the material, the instructor has the opportunity to spend extra time on a particular subject. By extension, the measures could be used to identify ideal times to implement interventions by instructors or researchers. As described by Yeager and colleagues (2013), changes to the classroom (e.g., interventions) could also be monitored and assessed by those same measures. Our suggestions here begin to highlight potential applications of pragmatic measurement, but it is important to move forward carefully and consider the trade-offs that may accompany users' adoption of the pragmatic perspective.

Table 1.
Current and Potential Uses Supported by Pragmatic Measurement

Current Uses	**Measure Type**
General Motivation Tracking (Example 1: This Chapter: Barron, Grays, & Hulleman, 2014)	Self-Report
Short-Term Repeated Measures (Example 2: This Chapter; Kosovich et al., 2017)	Self-Report
Intervention Development & Refinement (Example 3: This Chapter; Rosenzweig et al., in prep)	Self-Report, Open-Ended
Monitoring Potential Improvements to Classroom Activities (Yeager et al., 2012)	Self-Report

Intervention Fidelity Assessment (Wormington, Kosovich, Hulleman, in prep)	Rubric Coding & Profile Analysis
Experience Sampling	Self-Report, Open-Ended
Digital Learning Systems (Krumm et al., 2016)	
Potential Uses	**Measure Type**
Large-Sample Interviews	Interview, Open-Ended
Observational Checklists	Observation
Variance-Capturing Covariate Inventory (Broad list of items covering non-central but relevant sources of variance)	Any

Trade-offs and Drawbacks of Pragmatic Measurement

Unfortunately, the benefits mentioned above come without some of the methodological rigor present in more ideal measurement studies. First and foremost, electing to use brief measures precludes some of the most powerful statistical methods for conducting validation. For example, *confirmatory factor analyses* require at least three items per scale to test the statistical quality of items (Kline, 2011), and *item response theory* models may require even more for useful estimates (Embretson & Reise, 2000). Such techniques allow researchers to identify item bias where items tend to favor one particular group or context over another for reasons unrelated to the underlying construct. The inability to identify bias means that group differences could be true differences or simply differences in item interpretation. There are many instances where users may wish to compare groups on a particular measure. However, items containing substantial bias (i.e., characteristics that lead the measure to overestimate some groups' construct levels) may lead to inaccurate conclusions about the data. For example, does the measure miss critical aspects of an experience that would differentiate one group from another? Alternatively, is the measure focused on some critical aspects of an experience that would differentiate one group from another? Again this is a notable concern, and developers need to do whatever is possible to minimize such problems. In the case of pragmatic measurement, a different

measurement mode (e.g., qualitative interviews to support a Likert measure) may be able to identify some potential problems with a scale.

At a more basic level, using a single-item measure prevents the calculation of reliability coefficients (e.g., Cronbach's α). This is doubly problematic because, all else equal, scales can display lower reliability simply by including fewer items. Items that are unreliable and items that are biased can lead to false conclusions from the data. Furthermore, shortened scales are less likely to capture the full breadth of the construct under study (i.e., content validity), making it possible for overly brief measures to miss critical aspects of the construct. Such concerns are important and can lead to the attenuation, masking, or exaggeration of relationships or differences. Thus, it is important to consider both the evidence supporting (or not) the use of pragmatic measures as well as the consequences of inaccurate conclusions.

There is a nontrivial tradeoff between different factors related to conducting measurement. Pragmatic measures can be used to reduce intrusion into day-to-day routines, but doing so may result in lost information or inconsistent data. Adopting a pragmatic measurement approach can also reduce the likelihood that students become frustrated by the length of the assessment, but doing so may exclude variety of statistical tools for assessing the measures quality. The result is a need to find balance between the various considerations..

Moving Forward

Although we have found many instances of research that we would classify as using pragmatic measurement, there is no cohesive framework or body of literature on the practice. Moving forward, we are working to consolidate the existing research and to identify important considerations that need to be addressed more directly. For now, we simply advocate for identifying situational constraints in addition to the typical uses of the measure. The most critical

endeavor to undertake is to develop a specific method for considering the pros and cons of different decisions users may make. The method or framework then needs to be applied to different types of measurement (e.g., interviews, observations). The mode of measurement selected by researchers and practitioners has implications for what questions can be answered, what implementation methods are used to collect the data, and what conclusions can be drawn from that data.

Self-report. Self-report measures have the benefit of collecting data about individuals' personal experiences. Quantitative self-report measures (e.g., Likert-type scales, survey ratings) are able to compile data on a broad array of constructs from large groups of individuals with relative efficiency. Such measures do not allow users to probe answers for greater understanding, may fail to capture actual thought processes, may suffer from wording effects, and may frustrate participants with tedium (Bowman, 2010; Duckworth & Yeager, 2015; Paulhus, 1984; Schwarz & Oyserman, 2001). In contrast, more qualitative self-report measures (e.g., open-ended survey questions, interviews, focus groups) are able to delve into participant responses and response processes, and allow researchers to establish a rapport with the participants. Developing pragmatic self-report measures may be able to strike a balance between the benefits and drawbacks of these different approaches.

Observation. Another set of methods that can compensate for or complement the shortcomings of self-report measures are observational measures (Renninger & Bachrach, 2015), which vary in how flexible or structured they are. Observations (e.g., note-taking, video, artifact analysis) are desirable because they avoid the subjectivity or desirable responding in self-report, and they can avoid participant fatigue. However, observational measures can be less desirable in that they are at risk for observer biases, require rigorous observer training (Bill and Melinda

Gates Foundation, 2013), require more time for individuals to conduct the observations, and (in the case of less-structured observations) require more intense effort to analyze (Fredricks & McColsky, 2012). More structured observational measures (e.g., rubrics, checklists) can reduce the data collection and analysis burden but also may not be as readily able to capture unique occurrences or unexpected phenomenon.

Non-Human Observation. A somewhat different type of observation involves non-human observational tools, which can provide detailed data that depart substantially from the previous types of measures (e.g., Krumm et al., 2016). Non-human observational tools like eye-tracking and response times can be collected with little to no burden on the part of the participants. More complex measures like fMRI are more intrusive but may provide much more comprehensive information. Ultimately these types of measures may prove useful but are also more removed from the direct psychological experiences of individuals.

Different modes of measurement have their own benefits and drawbacks, but all measurement requires validation. An unreliable measure can hide relationships between constructs or produce spurious ones. An interviewer who has not been trained on a protocol may inadvertently influence the participant to respond in a desirable way. Thus, ensuring the technical integrity of measures is indispensable in conducting measurement. However, factors surrounding the measurement (e.g., limited data collection time or participant fatigue) could undermine validity in spite of excellent technical specifications (e.g., Wise & DeMars, 2005). Thus, potential implementation obstacles such as the time available to complete a measure need to be considered as the measures are being developed.

This chapter focused on self-reported quantitative data because the bulk of our research uses this type of measure. However, assessing pragmatic measurement's applicability to other

quantitative approaches is important. Even more so is the need to examine pragmatic approaches to qualitative data. Miniature interviews or highly specific lines of questioning could provide a wealth of information for quantitative and qualitative researchers alike. Developing methods that maintain high quality qualitative design for applied situations may be a significant hurdle.

In addition to the many suggestions we put forth in this chapter, we expect that many other viable ideas are simply not-yet conceived. There are many instances in which pragmatic measurement could be used to supplement or refine existing projects. For example, collecting measures to serve as covariates is an important but secondary concern in many studies. Thus, if we can develop pragmatic covariate measures that effectively serve as proxies for longer measures, we can either measure additional covariates of interest, or simply reduce the effort required from participants. Alternatively, we can increase the frequency of mixed-methods research by providing pragmatic supplemental measures for quantitative or qualitative researchers.

We end our discussion with a cautionary note. Pragmatic measurement is meant to foster innovation in research through flexible measurement practices. By pushing users to think more deeply about their purpose for measuring a construct, and to think about evidence that supports the use of a measure for those purposes, we can push the bounds of knowledge in new and interesting ways. However, users must be wary of carelessly adopting the first convenient set of items available. Our push for pragmatic measurement is in no way an endorsement for poor quality measurement work, nor is it a free pass for the widespread usage of brief or single-item measures. Every decision made about a scale introduces a trade-off between the technical and the pragmatic aspects of measurement and research. On the one hand, shortening a scale has the potential to drastically reduce its statistical integrity or the scope of conclusions that can be

drawn. On the other hand, failing to take situational factors into consideration may invalidate the data collected in spite of previous technical quality. The pragmatic perspective does not eliminate these trade-offs—instead it brings them to the forefront of validation.

References

AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: D. Van Nostrand Company, Inc.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, *41*(3), 586–598. http://doi.org/10.1037/0022-3514.41.3.586

Barron, K. E., Grays, M., & Hulleman, C. S. (2014). Assessing motivation in general education. In *Poster presented at the annual meeting of the American Educational Research Association, Philadelphia, PA.*

Barron, K. E., & Hulleman, C. S. (2015). Expectancy-Value-Cost Model of Motivation. In *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 8, pp. 503–509). Elsevier. http://doi.org/10.1016/B978-0-08-097086-8.26099-6

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment*, *17*(2–3), 62–87. http://doi.org/10.1080/10627197.2012.715014

Bill and Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching*. Retrieved from http://www.edweek.org/media/17teach-met1.pdf

Bowman, N. A. (2010). Can 1st-year college students accurately report their learning and development? *American Educational Research Journal*, *47*(2), 466–496. http://doi.org/10.3102/0002831209353595

Byrne, B. M. (2012). *Structural equation modeling with Mplus : basic concepts, applications, and programming*. *Multivariate applications series*.

Cronbach, Lee, J., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. *Educational Researcher*, *44*(4), 237–251. http://doi.org/10.3102/0013189X15584327

Eccles, J. S. (1983). Expectancies, Values, and Academic Behaviors. In J. T. (University of T. at A. Spence (Ed.), *Achievement-Related Motives and Behaviors* (pp. 119–146).

Eccles, J. S., Adler, T., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, Values, and Academic Behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco, CA: W. H. Freeman.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Fredricks, J. A., & McColsky, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). New York, NY: Springer. http://doi.org/http://dx.doi.org/10.1007/978-1-4614-2018- 7_37

Gaspard, H., Dicke, A., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*(9), 1226–1240. http://doi.org/10.1037/dev0000028

Goldstein, J., & Flake, J. K. (2016). Towards a framework for the validation of early childhood assessment systems. *Educational Assessment, Evaluation and Accountability*, *28*(3), 273–293. http://doi.org/10.1007/s11092-015-9231-8

Harackiewicz, J. M., Canning, E. a., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology*, *106*(2), 375–389. http://doi.org/10.1037/a0034679

Hektner, J. M., & Csikszentmihalyi, M. (1996). A longitudinal exploration of flow and intrinsic motivation in adolescents. In *Annual Meeting of the American Educational Research Association*.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. http://doi.org/10.1207/s15326985ep4102_4

Holstermann, N., Ainley, M., Grube, D., Roick, T., & Bögeholz, S. (2012). The specific relationship between disgust and interest: Relevance during biology class dissections and gender differences. *Learning and Instruction*, *22*(3), 185–192. http://doi.org/10.1016/j.learninstruc.2011.10.005

Hulleman, C. S. (2007). *The role of utility value in the development of interest and achieveement*. University of Wisconsin-Madison. Retrieved from http://files.eric.ed.gov/fulltext/ED498264.pdf

Hulleman, C. S., Barron, K. E., Kosovich, J. J., & Lazowski, R. A. (2016). Expectancy-value models of achievement motivation in education. In A. A. Lipnevich, F. Preckel, & R. D. Robers (Eds.), *Psychosocial skills and school systems in the Twenty-First century: Theory, research, and applications.* (pp. 241–278). Springer. http://doi.org/10.1007/978-3-319-

28606-8

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*.

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410–1412. http://doi.org/10.1126/science.1177067

Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2016). Making connections: Replicating and extending the utility value intervention in the classroom. *Journal of Educational Psychology*. http://doi.org/10.1037/edu0000146

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, *136*(3), 422–49. http://doi.org/10.1037/a0018947

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509–527. http://doi.org/10.1111/1467-8624.00421

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. http://doi.org/10.1111/jedm.12000

Kline, R. B. (2011). *Principles and practice of structural equation modeling (3rd ed.)*. *Principles and practice of structural equation modeling (3rd ed.).*

Kosovich, J. J. (2017). *Pragmatic measurement for education science: A method-substance synergy of validation and motivation*. University of Virginia.

Kosovich, J. J., Flake, J. K., & Hulleman, C. S. (2017). Short-term motivation trajectories: A parallel process model of expectancy-value. *Contemporary Educational Psychology*. http://doi.org/10.1016/j.cedpsych.2017.01.004

Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence*, *35*(5–6), 790–816. http://doi.org/10.1177/0272431614556890

Kosovich, J. J., Hulleman, C. S., & Flake, J. K. (2017). Practical measurement: An argument-based approach to exploring alternative psychometric validity evidence. In *Society for Research on Educational Effectiveness*. Washington, D.C.

Krumm, A. E., Beattie, R., Takahashi, S., D'Angelo, C., Feng, M., & Cheng, B. (2016). Practical measurement and productive persistence: Strategies for using digital learning system data to drive improvement. *Journal of Learning Analytics*, *3*(2), 116–138. http://doi.org/10.18608/jla.2016.32.6

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, *86*(2), 602–640. http://doi.org/10.3102/0034654315617832

Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. (1944). Level of aspiration. In J. M. Hunt (Ed.), *Personality and the behavior disorders* (pp. 333–378). New York: Ronal Press. http://doi.org/10.1037/10319-006

Messick, S. (ETS). (1989). Meaning and values in test validation: The science and ethics of

assessment. *Educational Researcher*, *18*(2), 5–11.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598–609. http://doi.org/10.1037//0022-3514.46.3.598

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667–686. http://doi.org/10.1037/0022-0663.95.4.667

Raynor, J. O. (1982). Future orientation, self-evaluation, and achievement motivation: Use of an expectancy x value theory of personality functioning and change. In N. T. Feather (Ed.), *Expectations and actions: Expectancy-value models in psychology* (pp. 97–124). Hillsdale, NJ: Erlbaum.

Renninger, K. A., & Bachrach, J. E. (2015). Studying triggers for interest and engagement using observational methods. *Educational Psychologist*, *50*(March), 58–69. http://doi.org/10.1080/00461520.2014.999920

Renninger, K. A., & Hidi, S. (2016). *The power of interest for motivation and engagement*. Routledge. Retrieved from http://works.swarthmore.edu/fac-education/91/

Rotgans, J. I., & Schmidt, H. G. (2011). The role of teachers in facilitating situational interest in an active-learning classroom. *Teaching and Teacher Education*, *27*(1), 37–42. http://doi.org/10.1016/j.tate.2010.06.025

Schmeiser, C. B., & Welch, C. J. (2006). Test Development. In R. L. Brennan (Ed.), *Educational Measurment* (4th ed., pp. 307–353). Westport, CT: Praeger Publishers.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22*(2),

127–160. http://doi.org/10.1177/109821400102200202

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, *74*(1), 107–120. http://doi.org/10.1007/s11336-008-9101-0

Traub, R. E., & Rowley, G. L. (1991). Understanding Reliability. *Educational Measurement: Issues and Practice*, *10*(1), 37–45. http://doi.org/10.1111/j.1745-3992.1991.tb00183.x

Vroom, V. H. (1964). *Work and motivation*. *Classic readings in organizational behavior*. New York: John Wiley & Sons, Inc.

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

Yeager, D. S., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). Practical measurement. *Palo Alto, CA: Carnegie Foundation for the Advancement of Teaching*, *78712*.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*(2), 267–301. http://doi.org/10.3102/0034654311405999

Appendix A

General Education Survey

We are interested in how students view [the university]'s general education program, *The Human Community*. This survey assesses your attitudes toward general education coursework at [the university]. For each section below, think about the courses you will take or have already taken to fulfill your general education requirements. Respond to each item using the following scale:

1	2	3	4	5	6
Strongly disagree	**Disagree**	**Slightly disagree**	**Slightly agree**	**Agree**	**Strongly agree**

Area One: Skills for the 21st Century

Area One is the cornerstone of The Human Community at [the university] in that students develop skills that facilitate inquiring, learning, thinking and communicating in their personal, academic, and civic lives.

Students at [the university] typically fulfill their Area One requirement by taking coursework in the areas of **critical thinking**, **oral communication**, and **writing**.

1. Area One courses interest me.
2. The topics and skills taught in Area One courses are important to me.
3. I can do well in my Area One courses.

Area Two: Arts and Humanities

Area Two shows students what it means to live lives enriched by reflection, imagination, and creativity. It does so by offering each individual a multidisciplinary experience within the arts and humanities, those areas of endeavor that humans have long valued for their intrinsic worth and that invite a deeper appreciation of the human experience.

Students at [the university] typically fulfill their Area Two requirement by taking coursework in areas such as **history**, **art**, **music**, **theatre**, and **literature**.

4. Area Two courses interest me.

5. The topics and skills taught in Area Two courses are important to me.

6. I can do well in my Area Two courses.

Area Three: The Natural World

Area Three provides students with the opportunity to develop problem-solving skills in science and mathematics at the college level. Students will be introduced to a substantial body of scientific facts, concepts, models, and theories and will also gain experience in using basic mathematics to obtain knowledge about the natural world.

Students at [the university] typically fulfill their Area Three requirement by taking coursework in areas such as **mathematics**, **chemistry**, **physics**, **astronomy**, **biology**, and **geology**.

7. Area Three courses interest me.

8. The topics and skills taught in Area Three courses are important to me.

9. I can do well in my Area Three courses.

Area Four: Social and Cultural Processes

Area Four helps students become critical thinkers about their own societies and the larger global community. These courses examine the key social and cultural processes and structures that shape the human experience.

Students at [the university] typically fulfill their Area Four requirement by taking coursework in areas such as **economics**, **history**, **political science**, **anthropology**, and **sociology**.

10. Area Four courses interest me.

11. The topics and skills taught in Area Four courses are important to me.

12. I can do well in my Area Four courses.

Area Five: Individuals in the Human Community

Area Five helps students learn about themselves as individuals and as members of different communities. Students explore how individuals develop and function in the social, psychological, emotional, physical, and spiritual dimensions.

Students at [the university] typically fulfill their Area Five requirement by taking coursework in areas such as **health**, **kinesiology**, **psychology**, and **sociology**.

13. Area Five courses interest me.

14. The topics and skills taught in Area Five courses are important to me.

15. I can do well in my Area Five courses.

Now that you have indicated your attitudes toward general education, we would like to know about your attitudes toward your major. Respond to the next three items using the scale below. *If you have not yet declared a major, select '7' as your response.*

1	**2**	**3**	**4**	**5**	**6**	**7**
Strongly disagree	**Disagree**	**Slightly disagree**	**Slightly agree**	**Agree**	**Strongly agree**	**I haven't declared a major yet.**

16. My major courses interest me.
17. The topics and skills taught in my major courses are important to me.
18. I can do well in my major courses.