5-4-2019

Developing a Multi-Dimensional Measure of Growth Mindset for School Improvement

Thomas Hartka

Developing a Multi-Dimensional Measure of Growth Mindset for School Improvement

Thomas A. Hartka

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

August 2019

FACULTY COMMITTEE:

Committee Chair: Dr. Kenneth Barron

Committee Members/ Readers:
Dr. Jeanne Horst
Dr. Keston Fulcher

To my parents, Marti and John.

Acknowledgements

I would be remiss without first giving a huge thanks to all of those who contributed to this process. There were many individuals that in large or small ways influenced the direction of this thesis or helped keep me sane. First and foremost, I would like to acknowledge Dr. Kenn Barron, my advisor and mentor. Kenn went above and beyond to make sure that I was able to receive timely feedback on my various drafts and of course guided me through each step of the process. Second, I would like to thank my committee members Drs. Keston Fulcher and Jeanne Horst. Their feedback and guidance on both the direction of the project and on the writing were so helpful.

Next, I would like to acknowledge the 6th grade math team Allison Yoder, Michelle Swartley, and Jim Raisch. It's no exaggeration to say that without their partnership and support none of this would have been possible. Their practical expertise and insight were invaluable.

I would also like to thank all those in my psychological sciences cohort and fellow GAs in the Center for Assessment and Research Studies. They helped keep me sane through the process as we shared thesis stories and helped point me in the right direction for credible sources.

Lastly, I have to thank my wife Blakeney. She lent a patient ear, tolerated my long nights, and put bread on the table during my Master's study.

Table of Contents

List of Tables

List of Figures

Abstract

The goal of the present study was to create a multi-dimensional growth mindset (MGM) measure. The purpose of the measure was to serve as an indicator of improvement for a team of 6th grade Math teachers in a local Middle School. These teachers noted that while their students were showing stronger self-reported growth mindset beliefs following interventions, they were not consistently displaying improvement in growth mindset behaviors. Following deeper discussions with this team of teachers and review of the growth mindset literature, six dimensions of growth mindset were identified: (1) intelligence belief, (2) effort, (3) persistence, (4) mistakes, (5) challenge, and (6) learning strategy. Development of a measure for these dimensions was driven by Benson's (1998) strong program of construct validity. The substantive stage, in which the researcher focuses on gathering theory to support the construct, was further driven by the seven steps of scale development outlined in Gehlbach and Brinkworth (2011).

With the measure created, a pilot study with the students of the partner teachers was conducted to gather evidence for the structural and external stages of Benson's (1998) program of construct validity. The structural stage is focused on evidence supporting how items are inter-related and related to the construct. An exploratory factor analysis revealed a six factor structure to the observed item scores. The emergent six factor structure was only moderately aligned to the theorized dimensions. The external stage is focused on evidence supporting how the measure is related to other constructs. Multiple regression models with the theorized dimensions and emergent six factors predicting English and Math grades and SOL scores, revealed that in general the classic

intelligence belief items were consistently the strongest predictors of educational outcomes.

Suggested next steps include further research in the substantive phase supporting how the proposed constructs are similar or different, improving the items on the MGM, and testing the measure with different populations and contexts.

Developing a Multi-Dimensional Measure of Growth Mindset for School Improvement

The goal of this study was to create and provide initial validity evidence for a new multi-dimensional measure of growth mindset. The development of this measure was situated in the context of an improvement project being conducted by a team of 6th grade teachers in a local middle school, which was itself situated in a larger body of improvement work being conducted at the same local middle school over the last several years (Barron & Hulleman, 2013; Barron, Hulleman, Hartka, & Inouye, 2017; Barron, Hulleman, Hartka, & Inouye, under review).

To provide more context for the current improvement project, I will begin by introducing a new model for school improvement proposed by the Carnegie Foundation for the Advancement of Teaching (Bryk, Gomez, Grunow, & Le Mahieu, 2015), as well as past improvement work at the local middle school that led to the present study. Then I will transition to the specific improvement project that motivated the need to develop a new growth mindset measure for the 6th grade math team. Although the new measure is being created to support a local improvement project for 6th grade teachers, the development of the measure has wider implications for theory, research, and practice, so careful attention will be paid to the collection of validity evidence. Benson's (1999) three stages of construct validation (substantive, structural, and external) will be used to guide and evaluate validity evidence for the new multi-dimensional measure.

Introduction

Background

Several years ago, Drs. Kenneth Barron and Chris Hulleman were invited to give a professional development workshop on motivation at a local middle school. They gave the session, but rather than just giving the session and ending the relationship, they decided to form a researcher-practitioner partnership with the school. This partnership included providing on-going professional development workshops and working with teams of teachers to develop and test interventions to be used in their classrooms to improve student motivation.

Soon after this partnership began, Drs. Barron and Hulleman joined a national improvement community led by the Carnegie Foundation for the Advancement of Teaching engaging in similar work. This community was called the Student Agency Improvement Community. The name highlights three elements central to community on which I will elaborate.

First, *student agency* captured the central goal of the community, which was to improve students' overall motivation and behaviors for learning. In particular, four key psychological areas were identified to foster student agency based on a comprehensive review of social psychological literature. These areas were growth mindset (students believe in themselves as learners), value (students find reason and purpose for learning), belongingness (students feel that they belong in the learning context), and learning strategies (students use effective learning strategies). Second, *improvement* denotes the method on how the community would conduct research, which was through the use of improvement science. Third, *community* highlights working collaboratively to accelerate learning faster than what one would typically learn working alone.

Improvement Science

At the core of the Student Agency Improvement Community was adopting the improvement science methodology. The Carnegie Foundation conceptualized six core principles of improvement (Bryk et al., 2015; see Table 1). Although the principles will be discussed in a particular order, it does not imply that the principles must be followed in that order. Improvement science is an iterative and cyclical process with various principles informing the others.

Table 1

The six core principles of improvement science (Bryk et al., 2015).

1) Make the work problem-specific and user-centered.
2) Variation in performance is the problem to solve.
3) See the system that produces the current outcomes.
4) We cannot improve at scale what we cannot measure.
5) Use disciplined inquiry to guide improvement.
6) Accelerate learning through Networked Improvement Communities.

The first principle of improvement is “Make the work problem-specific and user-centered.” The focus of this principle is to ground the work in an applied problem driven by the practitioner, rather than being motivated by a specific research question or theoretical problem driven by a researcher.

The second principle is “Variation in performance is the problem to solve.” Rather than focusing generally on what works, improvement is focused on determining what works, for whom, and under what conditions. This means the goal of improvement is not to find mean level change (shifting the normal distribution), but to produce a highly skewed distribution because changes have been introduced that benefit all (e.g., all students, all classrooms in a school, or all schools in a district). In order to accomplish this, it is important to identify both individual and situational factors where the system is not working yet.

The third principle is “See the system that produces the current outcomes.” A quote from a senior fellow of the Institute for Healthcare Improvement, Paul Batalden, summarizes this principle well, “Every system is perfectly designed to get the results that it gets.” The goal of this principle is to better understand the system surrounding the identified problem. To facilitate this learning, it is important to involve key stakeholders to gain multiple perspectives. Without adequate knowledge of the system, it can be difficult to bring about improvement, or to understand why something is not improving. The knowledge gained from studying the system is then used to guide the creation of a *working theory of improvement*. A working theory of improvement articulates the current understanding of what, where, and how changes might be introduced to help improve the system.

It is important not to jump to solutions too soon, which is a problem Bryk et al. (2015) refer to as “solutionitis.” Instead the focus of these first three principles is to make sure careful analysis has been done to fully understand the problem. Often the first “problem” identified is a symptom of a deeper embedded problem. As mentioned earlier, these principles should not be thought of as sequential steps. In this case, principles #1, #2, and #3 together create a feedback loop to help identify the root problem.

The fourth principle of improvement science is “We cannot improve at scale what we cannot measure.” To determine whether a change is an improvement, a measure needs to be tied to it. As discussed in principle #3, improvement science recognizes that improvement and changes do not exist in a silo, instead they are surrounded by a greater system. For this reason, the measurement should attempt to reflect that. Improvement science thus encourages the use of a *family of measures* rather than a single measure to assess improvement. For example, measures may be used to track (1) achievement of the overall aim of the improvement project, (2) intermediary outcomes, (3) efficacy of individual change ideas, or (4) unintended change in other areas of the system (i.e., balance measures). When considered together, a family of measures can gauge not only the efficacy of an improvement intervention, but also can capture the impact of the intervention in the greater context of the system. The fourth principle will become the specific focus of the current thesis, which is to develop a new measure that will be used as part of the family of measures for teachers in their classrooms.

The fifth principle is "Use disciplined inquiry to guide improvement." All improvement work is guided by three questions:

1. *What specifically are we trying to accomplish?*
2. *What change might we introduce, and why?*
3. *How will we know that a change is actually an improvement?*

These three questions mimic a small-scale research cycle where hypotheses for a change are made explicit, the hypotheses are tested against evidence, and ideas are modified based on what is learned and tested again. The way this plays out in improvement is through rapid testing cycles of <u>P</u>lan, <u>D</u>o, <u>S</u>tudy, <u>A</u>ct (PDSA). The goal is to learn quickly by doing, and to fail fast initially at smaller scales in order to learn faster to improve faster. A quote from an anonymous source highlights this well "Failures are not the problem; that we fail to learn from them is." This is not to say that educators should seek to fail. Instead, the focus is on gathering as much information as possible before rolling out a large scale intervention. When rolling out a new intervention in a new context it is likely that intended plans won't work perfectly immediately. Instead of failing on a large scale, the idea is to determine the efficacy and issues on a small scale and then scale up.

The sixth principle is “Accelerate learning through networks.” The focus of this principle is that learning is accelerated in community as opposed to the typical research silos that researchers commonly operate in. Working in a community, with other researchers and practitioners, can provide many benefits. But the primary benefit for improvement science is the ability to share learnings from both the successes and failures of community members. As highlighted in principle #2, the goal of improvement is not mean level change, but improvement for all. One example of principle #6 in action is a Networked Improvement Community (Bryk et al., 2015), for example the Student Agency Improvement Community. Networked Improvement Communities are guided by four features. First, they focus on a common aim (Principle #1 – make the work problem-specific and user-centered). Second, they engage in careful analysis of the system and build a working theory of how to improve that system (Principle #3 – see the system that produces the current outcomes). Third, they utilize improvement science methodology to rapidly and systematically design, test, and scale up improvement ideas (Principles #4 – we cannot improve what we cannot measure and #5 – use disciplined inquiry to guide improvement). Fourth and finally, they accelerate the rate and spread of learning by working collaboratively (Principle #6 – accelerate learning through networks) to test and adapt ideas for different educational contexts and student populations (Principle #2 – variation in performance is the problem to solve).

Improvement Work at Harrisonburg City Public Schools

Improvement work at Harrisonburg City Public Schools began with a root cause analysis of the motivational problems faced by teachers in their classrooms (Principle #1 – make the work problem-specific and user-centered). A root cause analysis is a process designed to assist in identifying the underlying factors or causes of a problem. While there were a variety of motivational issues identified, partner teachers agreed that their greatest issue involved students giving up on themselves as learners. Specifically, the teachers noted that many of their students did not believe in their ability to learn and be successful in school. Next, teachers were introduced to motivation theory and interventions that could increase students' belief in themselves as learners. Teachers determined that attribution theory, and in particular implicit theories of intelligence were particularly aligned with the motivational challenges they were seeing in students. Carol Dweck (2006) described two opposing theories of intelligence: fixed mindset – the belief that intelligence is a fixed trait that cannot be increased – and growth mindset – the belief that intelligence is malleable and can grow.

Based on emerging intervention work on growth mindsets (Paunesku et al., 2015; Yeager et al., 2016), our team began designing a growth mindset intervention that would be appropriate for the middle school context and easily delivered as an app on hand-held tablets. Through the Student Agency Improvement Community (Principle #6 – accelerate learning through networks), we received initial materials for an intervention designed for community college students (Gripshover et al., 2017). The team then engaged in a rapid series of PDSA cycles (Principle #5 – use disciplined inquiry to guide improvement). The first series of cycles focused on translating the intervention so that it would be appropriate for younger, middle school students. We first focused on creating a short intervention that 8th grade students could complete within 15 to 20 minutes. After having initial success with the oldest students in middle school (8th grade students), we also tested with 5th grade students (Principle #2 – variation in performance is the problem to solve) to ensure the intervention could be used with all middle school students.

To determine the effectiveness of the intervention, we put together a family of measures (Principle #4 – we cannot improve at scale what we cannot measure). The primary measure was a growth mindset measure (Dweck, 1999) to determine whether students' beliefs in the malleability of intelligence changed from pre- to post-intervention. We also included other process measures to track the students' experience with the intervention, including total time taken and student satisfaction. Based on results of these measures, we made various adaptations to the intervention. For example, while on average students finished in approximately 17 minutes, many students were quicker and others slower (especially if they were English language learners or students with learning disabilities). So, we added a reflective activity to complete after the intervention, part of which included a drawing activity that students could work on while others finished the intervention. Students were asked to draw two pictures in the activity. One asked students to draw themselves attempting something challenging at school and the other asked students to draw what happens to their brains based on what they learned in the app. This led to the discovery that the drawing activity provided a rich and unique manipulation check to determine how much students had gotten from the intervention.

After the initial series of PDSAs, we focused on how to scale up the intervention. One unexpected finding from the initial testing phase revealed that students could easily be exposed to the same intervention again during this process. For example, a member of our teacher team shared a number of students in common with other teachers, and her students commented if they should do the activity again if they already had done it in another class. A parallel learning was shared by teachers about a recent anti-bullying program implemented by the school. In the first year of the program, each grade was presented the same materials. In subsequent years, no changes were made to the materials and students continued to receive the same materials year after year. As a result, many students disengaged when hearing the material repeatedly and many teachers disengaged from wanting to implement the program as a result of students' reactions. Therefore, the team carefully analyzed the system and sought input from various teachers and administration (Principle #3 – see the system) to determine where the intervention would best be situated. It was decided to use it in 5th grade to serve as an introduction to students coming into middle school (since our partner school district placed 5th-8th grades together for middle school).

With the app situated in 5th grade, the next step was for 6th grade to develop their own unique growth mindset interventions. We identified a series of videos that were created in a partnership between PERTS and Class Dojo called Big Ideas. PERTS (Project for Education Research that Scales) is an applied research center based in Stanford University founded by students of Carol Dweck. The Big Ideas video series taught students about various growth mindset concepts. The math teachers created activities to go along with these videos, and through another round of PDSA cycles, we found that the videos and activities were effective in shifting students' growth mindset on the same measure (Dweck, 1999) used to evaluate our growth mindset app.

The New Improvement Problem to Address

The following year, the 6th grade math team set an aspirational aim to have "100% of their 6th grade students adopt a growth mindset by the end of the school year." To help accomplish this aim, we set up a plan to implement the previously tested Class Dojo growth mindset videos intervention at the start of year. Then we created and implemented growth mindset “booster” activities that would be strategically placed throughout the school year (e.g., after winter break and before taking standardized tests). The teachers also modified their teaching strategies to align with growth mindset principles to help communicate and reinforce growth mindset thinking more on a day-to-day basis.

After several interventions, the teachers shared an observation with us. We had found that the vast majority of students were, on average, reporting stronger growth mindset beliefs. However, students were not consistently engaging in growth mindset behaviors (e.g., being willing to routinely challenge themselves). In order to take a more systematic approach to understanding the problem, we began an improvement study focused on this problem that students were not consistently displaying growth mindset behaviors.

The Improvement Study. To understand the problem and its potential underlying causes more deeply, we conducted another root cause analysis (Principle #3 – see the system that produces the current outcomes). The tool we used to conduct the analysis was a fishbone activity (see Figure 1).

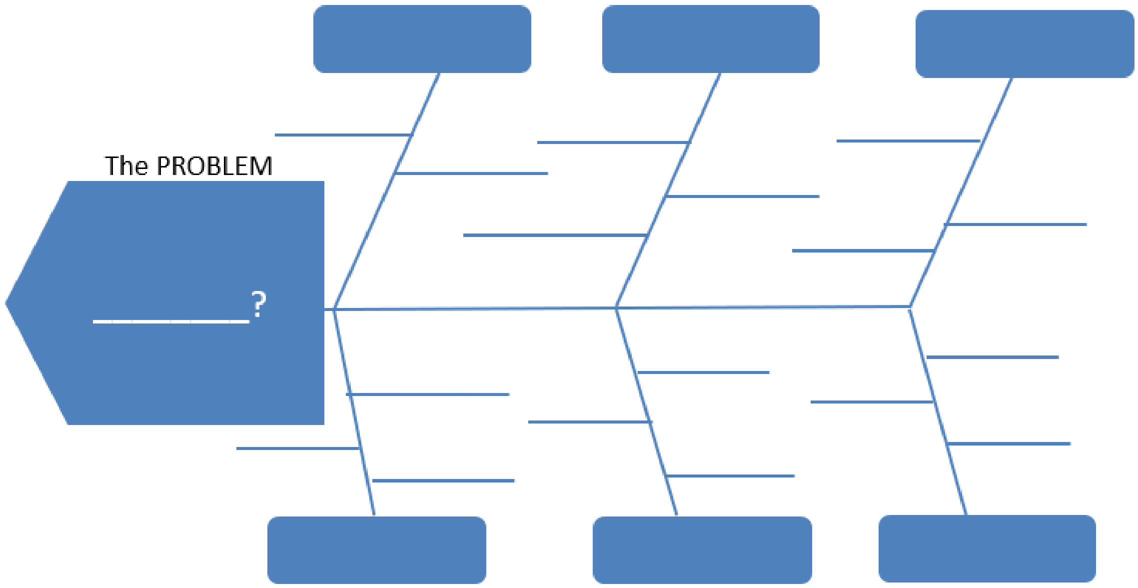

Figure 1. Image of a blank Fishbone Activity.

The fishbone activity starts with determining the problem statement. Specifically, we were trying to understand underlying factors or causes contributing to why current 6th grade students were not displaying growth mindset behaviors when they reported growth mindset beliefs. With the problem statement laid out, the next step is to brainstorm reasons the problem is occurring. After brainstorming possible causes individually, the group shares their causes and begins to arrange them together into categories. To help guide this activity, a visual is created (see Figure 1) that resembles the skeleton of a fish (aka, fishbones). The problem statement is written into the box to the far left that represents the head of the fish. Coming from this box is a central connecting line called the spine. Diverging from the spine are bones, which represent emerging groupings of possible causes. The boxes at the head of each bone allow those groupings to be labelled meaningfully into a theme. Below, is the fishbone created by the 6th grade math team and our researcher team (see Figure 2).

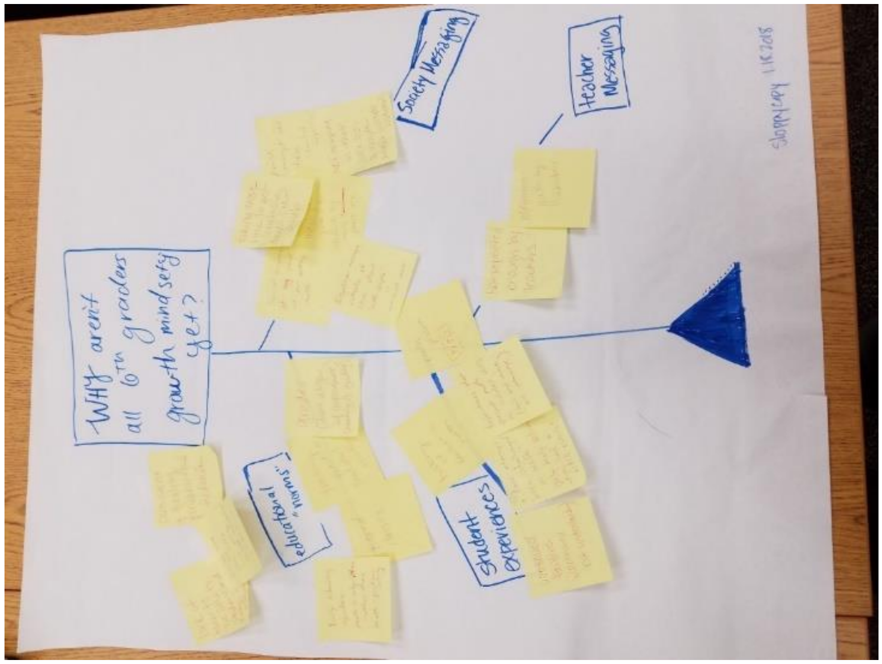

Figure 2. Image of the Fishbone Activity conducted with the 6^{th} grade math team.

Following the Fishbone Activity, we set about forming a working theory of improvement (Principle #3). In improvement, theory is purposefully referred to as *working* theory because as pieces of it are tested, they will be modified and refined to reflect new learnings. One common method for displaying the working theory of improvement is the Driver Diagram, which is what we used in this project (see Figure 3).

The first piece of the Driver Diagram is the aim statement (see Figure 3). The aim statement captures the ultimate improvement goal for a given project. A good aim statement will answer four key questions. What will be improved? By how much? By when? And For whom? The aim created by the math team was based on the goal they previously shared with us and edited according to their observation of inconsistent growth mindset behavior:

"All 6^{th} grade students will demonstrate consistent growth mindset beliefs and behaviors by the end of the school year."

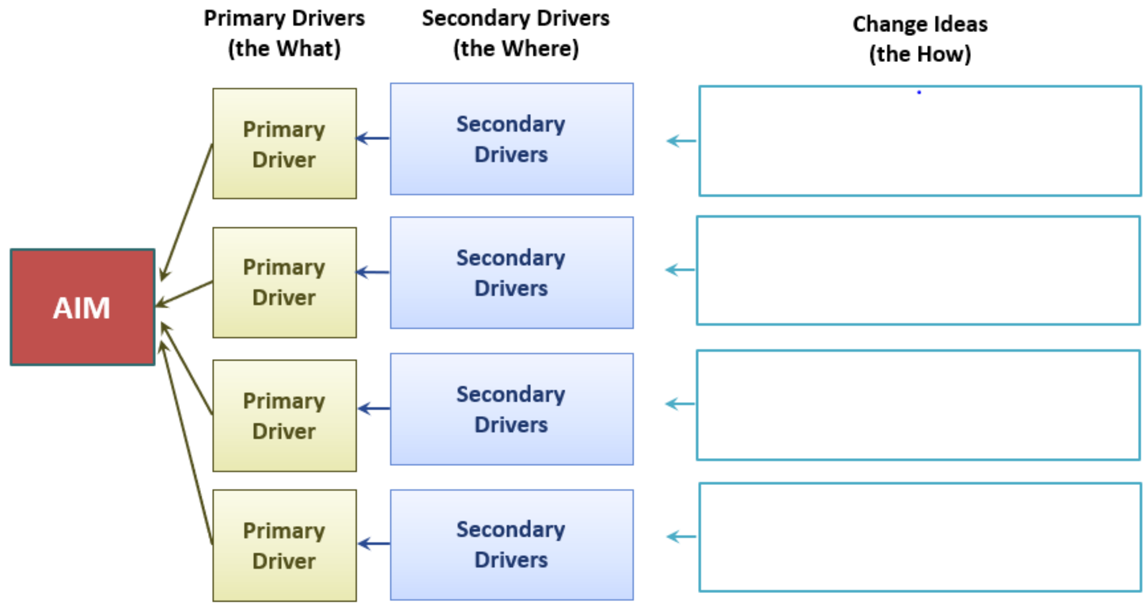

Figure 3. A full Driver Diagram including aim statement, primary drivers, secondary drivers, and change ideas.

The primary drivers are the next piece of the driver diagram (see Figure 3). The primary drivers represent *what* you need to do to reach the aim. In other words, the primary drivers are the major leverage points through which the aim will be accomplished. The primary drivers should be supported by the practical expertise of practitioners as well as theory. To identify the key growth mindset behaviors, we began with a review of the early growth mindset literature. The reason for focusing on early literature was that Dweck initially spent years building the foundation of growth and fixed mindsets from observing the behaviors of children when faced with educational challenges.

Early Growth Mindset Work. In one of her first studies (Dweck & Reppucci, 1973), Dweck followed up on the groundbreaking work of Seligman, Maier, and Greer (1968) who documented the phenomenon of learned helplessness in animals. Dweck demonstrated in a sample of 5th graders that, following failure, many students displayed a helpless response marked by deterioration of performance, while other students did not. Participating students also completed the Intellectual Achievement Responsibility (IAR; Crandall, Katkovsky, & Crandall, 1965) scale. The scale is made up of 34 forced choice items. Each item described a positive or negative achievement experience and then presented two response options. One attributed the cause to external circumstances, while the other attributed the cause to internal circumstances or behavior. Students displaying a helpless response tended to take less personal responsibility for outcomes (external attribution), whereas persistent students put more emphasis on the role of behavior and effort in determining outcomes (internal attribution).

Following this initial study, Dweck (1975) extended these results by using an attribution retraining method. Teachers identified 12 students that displayed extreme helpless responses and 10 additional students of equal ability who displayed persistence. These students initially completed a series of baseline measures, which included the IAR measure and an additional measure that had students choose between effort or ability attributions to failure situations. It was found that the two groups of students differed on both of the measures, corroborating Dweck and Reppucci (1973) and demonstrating that observed differences by teachers were reflected in differences on the associated measures. Following the completion of the measures, the helpless students completed an attribution retraining procedure or a control procedure. Students took the measures again, and students in the attribution retraining procedure showed a significant increase toward an emphasis of effort.

Following the two initial studies (Dweck & Reppucci, 1973; Dweck, 1975), the question remained why students displaying helpless responses performed worse than students displaying persistence when the two groups were theoretically of equal ability. To answer this question, Diener and Dweck (1978) conducted two studies. First, students were identified as helpless or mastery using the IAR scale and splitting students' data at the median. Next, students completed extensive training on a task. Following the training, students were presented with difficult versions of the task that would cause them to fail. The researchers then tracked students' use of strategies to arrive at a solution by asking students to engage in a think-aloud process (Lewis & Rieman, 1993) in which they vocalized their thoughts during the task.

In both studies, students with a helpless response style used ineffective strategies more often than mastery-oriented students. A deeper look into the results revealed that, over time, students with helpless responses showed a marked decrease in effective strategy use. On the first test problem, most students displayed the use of effective strategies, but by the last test problem, less than a third of students displaying helpless responses maintained those strategies. In comparison, mastery-oriented students did not show this decline, and in fact some of the students showed use of more sophisticated strategies in the face of difficulty.

When looking at students' verbalizations during the task, Diener and Dweck (1978) found a similar pattern of deviations over time. On the first test problem, helpless students began by vocalizing useful, on-task statements. However, after the onset of failure, differences in their vocalizations started to appear. Helpless students' verbalizations became characterized by attributions for failure ("I never did have a good memory"), a large number of solution-irrelevant statements (choosing brown colored tiles and saying "chocolate cake"), and negative affect ("This isn't fun anymore"). In contrast, verbalizations from mastery students did not show deterioration and were characterized by the absence of attributions for failure, the presence of self-monitoring (“I need to slow down and try to figure this out”), positive affect (“I love a challenge”), and positive prognosis (“I’ve almost got it now”).

Following this series of studies, Elliot and Dweck (1988) conducted a similar study but added achievement goal orientation. Achievement goal orientation refers to the particular disposition a student holds in regards to developing their abilities. At the time of this study, researchers were primarily focused on two specific goal orientations: (1) performance goals, which are characterized by individuals primarily seeking to maintain positive judgements of ability while avoiding negative judgements; and (2) learning goals, which are characterized by individuals primarily seeking to increase their ability or master new tasks. Specifically, the researchers hypothesized that helpless and mastery-oriented students pursued different goals, which would result in different patterns of behavior (see Figure 4 below). The study manipulated students’ perceptions of their ability (low or high) and goal (performance or learning). The task that students completed was the same task used in the Diener and Dweck (1978) study. After completing a set of problems, one experimenter gave feedback that the student had either low or high ability (random). Subsequently, a new experimenter presented students with two choices: a performance task (easy, medium, or hard) or a learning task (difficult and would make mistakes, but the student would learn something). The experimenter then made a certain goal salient: either a performance goal (students were being evaluated) or a learning goal (learning something here could help them in school).

The researchers then tracked students' (1) task choice, (2) strategies employed on the problems, and (3) verbalizations. Analysis of task choice revealed that when a learning goal was salient students more often chose the learning problem set, and when a performance goal was salient students chose the performance set. Differences were found for strategies employed between the low and high ability students in the performance goal condition with many of the low ability students deteriorating. In addition, it was found that students in the low ability, performance goal condition deteriorated over time, replicating the findings of Diener and Dweck (1978). Lastly, there were no differences between low and high ability students' verbalizations in the learning goal condition. In contrast, there was a significant difference between low and high ability students in the performance goal condition, where low ability students more frequently made attributions for failure, blaming their failure on some cause, than high ability students. These findings led to the initial theorizing of the wide range of beliefs and behaviors that may be impacted within these differing patterns of students' responses to failure.

Dweck and Legget (1988) later summarized findings from these early studies into two contrasting patterns of Affect-Behavior-Cognition (or ABCs for short): the maladaptive, helpless pattern and the adaptive, mastery-oriented pattern. The helpless response is characterized by negative affect, deterioration of performance in the face of difficulty, and an avoidance of challenge. On other hand, the mastery response is characterized by positive affect, challenge seeking, and the maintenance of effort. The authors then proposed a mechanism by which these behavior patterns are produced. First, individuals hold one of two contrasting implicit theories of intelligence (Dweck & Bempechat, 1983). Next, the implicit theory is proposed to influence what achievement goal is adopted (Dweck, Tenney, & Dinces, 1982). Ultimately, the achievement goal adopted will result in the overt pattern of adaptive or maladaptive behavior (Elliot & Dweck, 1988).

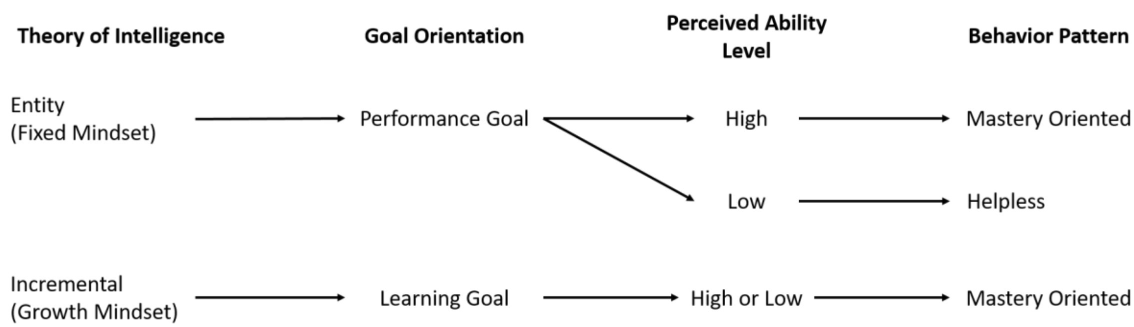

Figure 4. Pathway from theory of intelligence to overt behavior pattern. Adapted from Dweck and Leggett (1988).

Following the reporting of these adaptive and maladaptive attributional styles, researchers have studied how to promote the more adaptive growth mindset in education contexts (Aronson, Fried, & Good, 2002; Barron et al., under review; Blackwell, Trzesniewski, & Dweck, 2007; Paunesku et al., 2015; Yeager et al., 2016). However, a recent meta-analysis (Sisk, Burgoyne, Sun, Butler, & Macnamara, 2018) illustrated that while many studies have found a positive relationship between growth mindset and achievement, others have found no relationship or even a negative relationship between growth mindset and achievement. Wormington et al. (under review) showed a similar complex pattern of results. When testing the efficacy of a growth mindset intervention in community college, they found no difference in course grades between individuals in the growth mindset and control conditions. Those participating in the growth mindset condition were also asked to write short essays, which were later coded for themes. They found that individuals who wrote only about the growth mindset belief again fared no better than control students on academic performance. However, they found that individuals who discussed both growth mindset beliefs *and* specific behaviors (such as learning strategies) showed significantly higher academic performance than individuals in the control condition.

These studies support that a growth mindset intelligence belief alone is not enough to result in improved achievement. Furthermore, in a recent review article, Dweck and Yeager (2019) discussed the prevalence of “false growth mindsets.” A phenomenon where a practitioner does not fully grasp growth mindset and fails to adequately communicate it to students. They noted that, “some educators told their students that they could do anything but did not provide them with strategies, guidance, or information about resources for the accomplishment of this promise” (pg. 10).

Overall, this review of the literature suggests that growth mindset is more complicated than a unidimensional belief that intelligence can be improved. This is not to undermine the importance of the belief. Prior research highlighted here and work on growth mindset interventions highlight the importance of the intelligence belief. However, there appear to be several areas key to promoting growth mindset beliefs and behaviors.

The first of these areas to consider involves effort. Differences between students on effort were found in Dweck’s earliest articles (Dweck & Reppucci 1973; Dweck, 1975). The second area to consider involves persistence. As noted in Dweck (1975), a particular student group was identified as persistent. Furthermore, Diener and Dweck (1978) demonstrated that mastery-oriented students persevered and did not deteriorate in the face of difficulty, even when problems were unsolvable. The third area is challenge. Elliott and Dweck (1988) demonstrated that students adopting performance goals would avoid challenge, while students adopting a learning goal would seek challenge. The fourth area identified is mistakes. This entire line of research was built upon observations of how students responded to failure and making mistakes. Finally, the fifth area is learning strategies. As Wormington et al. (under review) and Dweck and Yeager (2019) discussed, if students do not have effective learning strategies and support structures they will not succeed. These five identified areas along with intelligence belief were adopted as six primary drivers (Figure 3) by which we would accomplish the aim of having students display more consistent growth mindset beliefs and behaviors.

Following the primary drivers, the next aspect to consider is *where* in the system the primary drivers will be situated, which are referred to as secondary drivers. In the fishbone activity, the math teachers identified four sources for why students may not be displaying growth mindset behaviors: messaging from teachers, their experiences in school, educational norms, and societal messaging (messaging they hear outside of school). These were areas where teachers felt students might be hearing/seeing messages counter to growth mindset. After reflecting on these growth mindset threats and discussing them further, we identified five secondary drivers:

1. Direct-to-student interventions: this would include activities like growth mindset interventions (references) where the activity can be delivered right to a student regardless of the context.
2. Changes to teaching strategies: the specific learning opportunities given to students and how they are presented can reinforce growth mindset concepts. For example, using a mastery grading system over a performance grading system.

3. Changes to school norms: this refers to the greater system in place for a school or school district. For example, procedures for punishing bad behavior, communicating to parents and students, communicating between administration and teachers. All of these systems in place in the school and school district contribute to the creation of the culture and norms.
4. Student-student interactions: this is focused on language. Specifically, how students are communicating to each other during class.
5. Parent-student interactions: similarly to #4, this is focused on the language that parents use to talk about school and learning with their students.

The last section of the Driver Diagram involves Change Ideas (see Figure 3). These are specific change ideas (aka, interventions) that describe *how* to improve a system to reach the aim. For example, a change idea that was implemented to help parent-child interactions was a set of questions attached to report cards for parents to ask their children. These questions were framed in a growth mindset way and focused on the process of learning (e.g., what class did the student feel most proud of, what was the most interesting thing the student learned) rather than the product (the grade). Following the logic of the driver diagram to accomplish the aim that “all 6^{th} grade students will demonstrate consistent growth mindset beliefs and behaviors by the end of the school year,” students will need to adopt a stronger growth mindset belief. To help foster better communication between parents and students, when report cards are sent home a set of questions will be accompanied to help guide a more productive discussion of the learning process (see Figure 5).

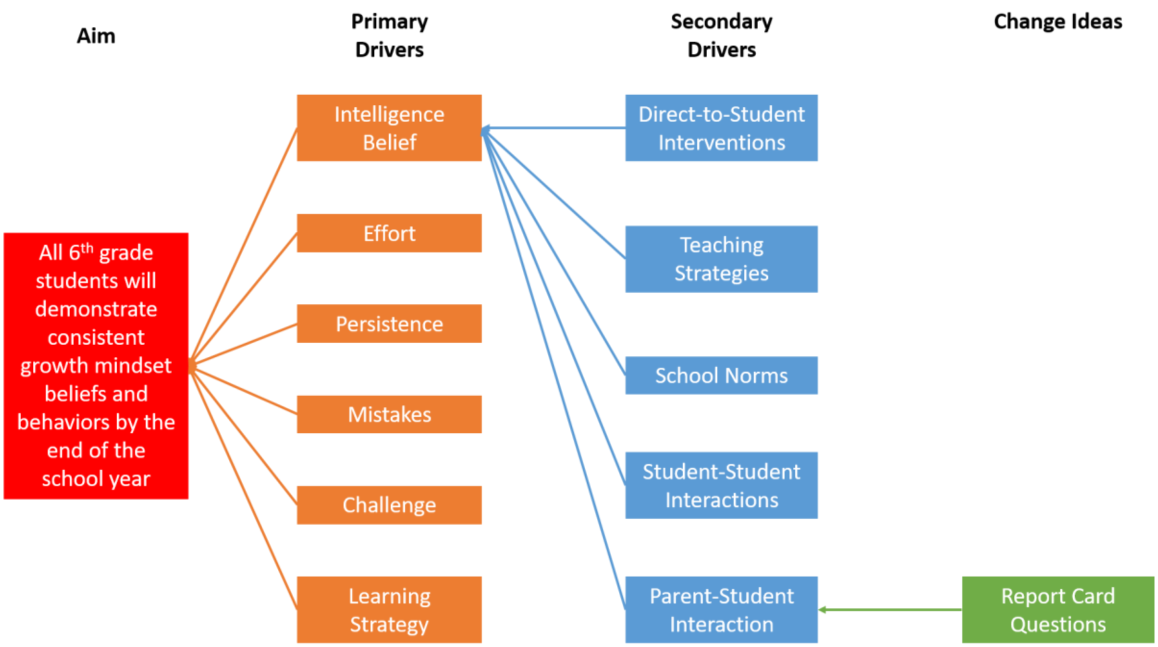

Figure 5. Completed 6th grade math driver diagram. For simplicity, the secondary drivers were only mapped to intelligence beliefs. In a full driver diagram, they would be mapped to each primary driver.

With a driver diagram in place, our next goal was to identify a family of measures that would inform us whether movement toward the aim was occurring (Principle #4). In improvement, a robust family of measures is very useful. First, there should be a measure in place for the aim. Then, there should also be proximal measures that will inform whether movement on the primary drivers is occurring. As the working theory of improvement suggests, if movement on the primary drivers is occurring there should also be movement on the aim. If this is not occurring, the working theory may need to be modified or refined. There is a quote in improvement often tied to the Driver Diagram: “Possibly wrong; definitely incomplete” (Bryk et al., 2015). This captures the spirit of continuous improvement, which in itself is a growth mindset oriented process. As failures occur, improvement takes the perspective that they are valuable and can inform the working theory. Lastly, there should be measures tied to the individual change ideas.

These measures will evaluate the efficacy of specific changes and interventions and their value within the overall improvement system. With the aim that students will display a growth mindset pattern of both beliefs and behaviors, a new measure was necessary to evaluate whether all the drivers of the growth mindset are being achieved.

The Present Study

The problem posed by 6^{th} grade team of math teachers suggested that the current measure of growth mindset being used, a shortened version of the Implicit Theories of Intelligence Scale (Dweck, 1999), was not sufficient. The purpose of the present study was to create a new Multi-dimensional Growth Mindset measure (MGM) that may be used in the family of measures for improvement work being conducted by the math team. Specifically, this overall measure will serve as the measure of the aim, and the subscales may be used individually as measures for the primary drivers. The study will also provide preliminary validity evidence for the measure. The framework that will be used to evaluate validity evidence is Benson's three stages of construct validation (Benson, 1998).

A construct represents a latent variable about which researchers or practitioners are interested in. Cronbach and Meehl (1955) define a construct as an "attribute of people, assumed to be reflected in test performance" (pg. 283). Because most constructs cannot be directly observed (i.e., latent), inferences must rely on scores from measures. For example, a common construct of interest in education is intelligence. It is not something that can be measured directly. Instead, we rely on performance assessments from measures such as the WAIS (Wechsler Adult Intelligence Scale; Wechsler, 2008) to make inferences about an individual's intelligence quotient (IQ). The process of construct validation provides supporting evidence that those inferences that are being made are adequate and appropriate (Messick, 1989).

Cronbach (1989) described two programs of construct validation, strong and weak. A strong program of construct validity depends on precise theory and is reflected in the *Standards* (AERA, 2014). In contrast, a weak program of validity stems from less fully articulated theory and construct definitions. Benson (1998) extended the strong program of construct validity by outlining a process consisting of three stages: substantive, structural, and external. The purpose of the *substantive stage* is to determine how a construct is defined, operationalized, and ultimately measured. During this phase, researchers engage in a review of theory and previous research to define the construct. The *structural stage* involves relating the items of the measure to the structure of the construct, and determining the extent to which items relate to one another and to the construct. This stage is typified by psychometric investigations into such properties as internal consistency and factor structure. Lastly, the external stage involves determining whether the measure of the construct of interest relates in expected ways with measures of other constructs. Researchers often gather such evidence as convergent, divergent, and predictive evidence during this stage.

The present study is focused on addressing the three stages of Benson (substantive, structural, external) in the creation of a new multi-dimensional growth mindset measure. The processes involved in identifying different theoretical dimensions of growth mindset along with writing items to create a new survey to measure each dimension are housed within the substantive stage. Following creation of the survey, a study was conducted from which statistical evidence was collected for the structural and external stages.

Substantive Stage: Construction of the New Measure

To conduct this stage, I used an approach outlined by Gehlbach and Brinkworth (2011). This process was created to bolster the collection of validity evidence from measures. It is composed of seven steps (see Table 2).

Table 2

Seven steps to gather validity evidence for a measure (Gehlbach & Brinkworth, 2011).

1)	Literature Review
2)	Interviews and Focus Groups
3)	Synthesizing the Literature Review with Interview/ Focus Group Data
4)	Developing Items
5)	Expert Validation
6)	Cognitive Pretesting
7)	Pilot Testing

Too often validity evidence is thought of as empirical data gathered after a measure has been created, but Gehlbach and Brinkworth (2011) highlight multiple steps to help front-load validity considerations focused on the substantive phase. By taking validity into consideration from the inception of a scale, it allows for various lines of validity evidence to be planned a priori. In addition, thought may be given to whether certain areas are of more concern than others. For example, certain areas may be of higher priority depending on the intended use of the scale. Thus, those areas can be targeted from the outset. As Kane (1992) puts it, “Validity evidence is most effective when it addresses the weakest parts… evidence that provides further support for a highly plausible assumption does not add much to the overall plausibility of the argument” (pg. 530).

Following Gehlbach and Brinkworth (2011), the first step in the process is the literature review. As our team began to consider the process of developing a Multi-dimensional Growth Mindset measure, I was interested how Growth Mindset measurement evolved over time (Table 3). Understanding how it progressed would give our team a better understanding of how Growth Mindset measurement evolved and the current thinking in theory. In addition, we thought it would help enlighten us on how to move forward with the creation of a new measure.

Growth Mindset Measurement History. As noted earlier, before Carol Dweck began her research on implicit theories of intelligence, she studied the observation that certain children in an education setting would exhibit a helpless response while others would continue persisting. Based on these observations, one of the first measures identified was the Intellectual Achievement Responsibility (IAR) Scale (Crandall et al., 1965), which consisted of 34 forced-choice items. Each item described a positive or negative achievement experience and was followed by two response options. One attributed the cause to external circumstances, while the other attributed the cause to internal circumstances or behavior. Of particular interest was a subset of 10 items that specifically dealt with attributions of failure to lack of effort. Dweck and Reppucci (1973) provided initial evidence for the use of the measure after it successfully differentiated students who were displaying a helpless response from those displaying a mastery-oriented response.

Table 3

History of measures used to assess Growth Mindset.

Measure	Study

1)	Intellectual Achievement Responsibility (IAR) Scale	Crandall, Katvosky, & Crandall (1965); Dweck & Reppucci (1973)
2)	Direct observation of students	Dweck (1975)
3)	Implicit Theory - IAR	Dweck & Bempechat (1983)
4)	3-item measure	Henderson & Dweck (1989)
5)	Implicit Theories of Intelligence Scale (ITIS)	Dweck (1999)
6)	Shortened ITIS	Various studies

Following the initial use of the IAR, Dweck (1975) provided further support for its use by connecting it to observations. Dweck had teachers, a principal, and a school psychologist select students who displayed the most extreme helpless response as well as students of equal ability who displayed a mastery response to evaluate known-groups validity. Once these groups were identified, they were compared on the IAR, and it was discovered that the pattern of responses was similar to those obtained by Dweck and Reppucci (1973) supporting known-groups validity. While providing strong evidence for the use of the scale, results also set up a precedent that the underlying beliefs being measured were connected to overt behaviors.

Subsequent studies (Diener & Dweck, 1978; 1980) continued to use the IAR scale to differentiate between helpless and mastery students. In 1983, Dweck and Bempechat conceptualized the entity and incremental theories of intelligence. With the success of the IAR, a revised version of the IAR focusing on implicit theories of intelligence was created. This scale presented students with contrasting statements, one endorsing the entity theory and the other endorsing the incremental theory. For example, one item presented students with the following choice:

a) *You can learn new things, but how smart you are stays pretty much the same.*

b) *Smartness is something you can increase as much as you want to.*

They soon discovered that students would begin selecting more incremental statements as they progressed through the measure (Dweck & Henderson, 1989). This indicated that the incremental statements were either highly compelling or perhaps also socially desirable. To eliminate the probable response bias, the items were shifted to a format where students were asked how much they agreed or disagreed with three entity theory statements:

1) *You have a certain amount of intelligence and you really can't do much to change it.*
2) *Your intelligence is something about you that you can't change very much.*
3) *You can learn new things, but you can't really change your basic intelligence*.

This form of the scale was used in studies moving forward (Dweck, Chiu, & Hong, 1995; Henderson & Dweck, 1989; Hong, Chiu, Dweck, Lin, & Wan, 1999). Further validity evidence for the three-item scale was provided in Dweck et al. (1995) supporting its use. Levy and Dweck (1997) provided validity evidence for an extended version of the scale with eight items, including four items oriented toward an entity theory of intelligence and four items oriented toward an incremental theory. The eight-item scale was later published for wider spread use by Dweck (1999) as the Implicit Theories of Intelligence Scale (ITIS). Even though a lengthier scale was created, which would help reduce measurement error, subsequent studies continued to use a shortened version of the scale (Aronson et al., 2002; Blackwell et al., 2007). Even recent studies have continued using a reduced version of the ITIS (Bettinger, Ludvigsen, Rege, Solli, & Yeager, 2016; Froelich, 2016; Romero, Master, Paunesku, Dweck, & Gross, 2014; Yeager et al., 2016). This may be due to a reason provided in Hong et al. (1999) who stated: “Only three items are included because the items are intended to have the same meaning, and continued repetition of the same idea becomes somewhat bizarre and tedious” (pg. 590). In our own research, we also have used a shortened version of the ITIS (Barron et al., 2017; Barron et al., under review).

In summary, the current version of the ITIS is often shortened when used in research studies in education contexts. The reason is that there is often limited time availability, and the full scale can often be tedious for students to fill out. Thus, the most common form is a two- or three-item scale that is focused on the entity theory of intelligence. However, as our teachers and the research has pointed out, this inward intelligence belief may not be enough to change outward behavior.

The second step of Gehlbach and Brinkworth (2011) is to explore the construct of interest through interviews and focus groups. I conducted over a dozen interviews with 6th grade students from the classrooms of the partner math teachers. The interviews were semi-structured, such that interviewers had a list of questions to work through within each of the identified dimensions (see Appendix A). However, if interesting topics came up, interviewers had the opportunity to probe for more information. We found that within effort beliefs, most students believed that, through effort, intelligence could grow. This provided support that the greater body of improvement work being conducted at the school was effective. Additionally, several students actually cited interventions conducted by 6th grade teachers’ improvement work for why students held that belief. Further, the results showed a greater range of beliefs within each of the other dimensions (challenge and mistakes). For example, within the challenge dimension some students believed challenges were a good thing, others felt that challenges were both good and bad, and other students didn't like challenge.

One of the major findings of the interviews was that students did not know what learning strategies meant. To facilitate some discussion when this happened, the interviewer would provide a definition and an example. However, students rarely provided any novel information beyond that provided by the interviewer.

The third step is to synthesize the literature review with qualitative data collected through interviews and focus groups. The findings from the interview support that intelligence beliefs and effort beliefs are closely tied together. The section on effort beliefs, overall, contained the most positive responses from students. With much of the improvement work at THMS focused on growth mindset, this finding reinforced that it has been effective. On the other hand, overwhelmingly students did not know what learning strategies were. This is concerning, because a recent finding from Wormington et al. (under review) found that acknowledgment of learning strategies was essential. They found that only students who wrote about both the growth mindset belief and learning strategies received significantly higher course grades. Conversely, students who did not write about any learning strategies did not do any better than the control students.

The interviews also revealed that there was a wider range of beliefs held by students within challenge and mistake dimensions, than other dimensions. This finding supports the importance of multiple dimensions. These dimensions have been cited as important to growth mindset, however, in the sample of students interviewed, having high effort beliefs did not also lead to having positive beliefs about challenge and mistakes. The findings from the interviews reinforced that growth mindset interventions are influencing students' effort beliefs, but may not be influencing their beliefs about challenge and mistakes.

Next, the fourth step is to construct the survey. Our primary consideration was the context in which the measure would be situated and how it would be used. We recognized the primary purpose of this survey was to serve as a practical improvement measure. Yeager and Bryk (2015) shared multiple considerations for the development of improvement measures. First, improvement studies require direct measurement of intermediary targets. Our earlier improvement processes revealed that six areas were identified as key areas that would support the development of growth mindset: (1) intelligence belief, (2) effort, (3) persistence, (4) challenge, (5) mistakes, and (6) learning strategies.

Second, practical measurement often presses toward greater specificity. The multiple dimensions of the measure will allow us and teachers to better understand how growth mindset interventions are working. Previously, we would only be able to determine how interventions performed with respect to intelligence beliefs. With a multi-dimensional measure, we ideally will be able to better understand what growth mindset beliefs and behaviors current interventions are addressing, or not addressing yet. This in turn can inform how current interventions may be modified or how new interventions can be created.

Third, increased validity can be achieved when items on the measure reference that which you are hoping to change. In this case, that would be the primary drivers of the driver diagram (see Figure 5). Furthermore, within each of the dimensions I also considered what type of questions to ask. Looking back at the early literature, we found compelling some of the early writing of Dweck and colleagues (Dweck & Leggett, 1988; Henderson & Dweck, 1990). They used a framework of Affect-Behavior-Cognition (ABC) to describe growth versus fixed mindset patterns of response. Using this framework of ABC to guide the development of questions gets closer to students' actual experience of the dimensions as opposed to an entirely belief (or cognitively) oriented measure. See Appendix B for the full scale. For example, below are items that were proposed for the effort dimension:

1) *If I have to work hard in school, it makes me feel bad about myself. (Affect)*
2) *I often work hard in school. (Behavior)*
3) *When I have to work hard in school, it makes me think that I am not very smart. (Cognition)*

Fourth, measures need to be engineered to be embedded within the constraints of everyday school practice. A particular challenge was to create a practical improvement measure while simultaneously attempting to address six dimensions. I decided to honor the ABC framework and have one question addressing affect, behavior, and cognition for each dimension except intelligence belief, which will continue to be tested with the shortened version of the ITIS. This creates a target goal of 17 items for the proposed Multi-dimensional Growth Mindset (MGM) measure. While overall the measure is longer than Yeager and Bryk (2015) suggest, I also recognized that teachers may choose to only use one dimension of the overall measure to assess the efficacy of certain interventions. For example, to test an intervention designed to get students to approach challenge more, teachers could choose to only use the challenge subscale. This versatility has added practical utility as a number of our teachers already use quick exit ticket assessments of 2-4 items to measure key objectives they are interested in evaluating.

Once we created the plan for the constructing the survey, we engaged in several rounds of brainstorming and revising items. To facilitate the process, we also looked at measures related to our dimensions that may be relevant. For example, we looked at other growth mindset scales (Abd-El-Fattah & Yates, 2006; De Castella & Byrne, 2015; Faria & Fontaine, 1997), GRIT scales (Duckworth, Peterson, Matthews, & Kelly, 2007), and effort belief scales (Blackwell et al., 2007; Froelich et al., 2016).

After constructing the survey, the next step was to seek expert validation. We sent the survey back to our participating teachers to get their opinions on the survey. Specifically, we were interested in whether they felt the questions would give them actionable information and whether students would be able to understand the questions. From this feedback, we prioritized certain questions and changed the wording of others.

With feedback from the teachers on the survey, I veered from the steps in Gehlback and Brinkworth (2011). Rather than following up with cognitive testing, I had a timely opportunity to test the initial measure with the 6th grade students of our partner teachers. The purpose of this initial study was to gather preliminary evidence regarding how the dimensions, ABC framework, and individual items were operating by moving on to Benson's (1998) structural and external stages.

Structural and External Stages: Initial Study and Psychometric Evidence

Benson's (1998) structural stage shifts the focus from understanding the "substance" of the construct to examining psychometric evidence. The purpose is to provide statistical evidence that the items in the measure are inter-related in expected ways. The external stage is similarly looking at relationships, however instead of internal relationships, the focus is on external relationships. In other words, how are the items and constructs in the measure related to other constructs? In order to collect this information, a study was conducted with partner 6th grade math teachers.

Methods

Participants. Participants in this study were 200 6th grade students at Thomas Harrison Middle School, one of the two middle schools in the Harrisonburg City Public School system. Students in the sample were 50% female, 43% Hispanic, 40% White, and 13% Black. Additionally, 16% of students were of limited English proficiency. Parents were given the opportunity to opt their students out of the study and students were given the opportunity to assent to participation. As a result of these processes, 48 students were omitted from the final data set for a total of 152 students in the final sample. It is important to note that all students were still required to complete the survey, because their teachers were interested in using results for their own improvement purposes. If students chose not to have their responses included in the study, their responses were omitted and were not included in the analyses below.

The current participants were chosen for two reasons. First, this is the population the math teachers work with. The same population from which they observed that students were not consistently engaging in growth mindset behaviors. Second, developing this measure with 6th grade students (the youngest students in the middle school) helps ensure it will be appropriate, in terms of reading level, for use with all grades in middle school.

Measures

Multi-dimensional Growth Mindset. Only one measure was included in this study: the newly proposed Multi-dimensional Growth Mindset **(**MGM) measure (see Appendix B). The measure included 18 total items. Each of the newly proposed dimensions of effort, persistence, challenge, mistakes, and learning strategies was assessed with three items. Those three items were designed to capture affective, behavioral, and cognitive (ABC) components of each respective dimension. In addition, a two-item version of the ITIS used previously by the researchers for improvement work at HCPS (Barron et al., 2017; Barron et al., under review) to measure intelligence beliefs was included, as well as a one-item manipulation check asking students how honest they were with their responses. Students rated the extent they agreed or disagreed with each item using a six-point Likert scale (1 *Strongly disagree* – 6 *Strongly agree*). Items that were negatively worded were reverse scored and from here on will be marked with an indicator (**Reversed*) at the end of the item. With negatively oriented items reverse scored, this indicates that values closer to 6 reflect having better affective, behavioral, and cognitive reactions toward that dimension. For example, both intelligence belief items are framed negatively (fixed mindset oriented) where a score closer to 6 would indicate a fixed mindset. After reverse scoring, a score close to 6 indicates a strong growth mindset and a score close to 1 indicates a strong fixed mindset.

Achievement. Two measures of achievement were collected. First, final grades for the 6th grade students were obtained in English and Math. Each final grade represented students' overall performance in that particular core class for the entire 2018-

19 academic year. Grades were given to students as both a raw score (0-100) and as a letter grade (A, B, C, D, F). The present study used raw scores in analyses.

Second, standardized testing scores for 6th graders were also collected. In Virginia, standardized tests are referred to as the Standards of Learning (SOL). In 6th grade, there are two SOLs: Reading (English) and Math. Scores on these tests range from 200 to 600. Students are able to pass these tests at two levels of proficiency, Pass (500) and Pass Advanced (600). As with final grades, the present study only used students' raw scores on the SOLs.

By collecting both graded and standardized performance, I could compare and contrast how different dimensions of growth mindset would relate to the two end-of-the-year outcomes that our teachers were trying to improve in their students.

Procedure

The MGM survey was completed by students in the context of their math class in the spring of the 2018-19 school year. The primary researcher, Thomas Hartka, introduced the purpose of the study, informed students of their rights as participants, and allowed them to indicate if they would like to officially participate in the study. Students then completed the measure using Chromebook computers. During the survey, the researcher worked with the math teacher to provide read aloud support to students who needed it. In the current study, read aloud support was provided for two of the nine math classes. After all students completed the survey, the researcher thanked students for their participation.

Results

Results for the structural and external stages will be organized in separate sections below.

Structural Stage Analyses

Recall the structural stage of Benson's (1998) framework focuses on how scale items are inter-related and related to the proposed construct(s). To gather evidence for this stage of the validity process, I first conducted preliminary descriptive analyses on each item. Next, I conducted inter-item correlations followed by an exploratory factor analysis to determine how items were related to one another and how they were operating together. Last, internal consistency was estimated for the theorized subscales and emergent subscales from the factor analysis.

Descriptive Statistics. Descriptive statistics were run to examine item level performance (Table 4). On average, the most strongly endorsed item from students was the effort affect item (*If I have to work hard in school, it makes me feel bad about myself. *Reversed; M* = 5.08, *SD* = 1.11) and the least strongly endorsed item was the persistence affect item (*I get frustrated if something takes a long time to learn.*Reversed; M* = 3.31, *SD* = 1.55). Remember that because this item was reverse scored a lower score indicates higher frustration. When comparing the pattern of responses from subscale to subscale, intelligence belief items were the most strongly endorsed (*M* = 4.99, *SD* = 1.20) while challenge items were the least strongly endorsed (*M* = 4.09, *SD* = 1.26). Similarly, comparing the pattern of responses between affect, behavior, and cognition items revealed that behavior items were the most strongly endorsed (*M* = 4.68, *SD* = .78) while affect items were the least strongly endorsed (*M* = 4.09, *SD* = 1.02).

Table 4

Item level descriptive statistics for the MGM.

Item	*M*	*SD*	Range	Skewness	Kurtosis
(ib1) You have a certain amount of intelligence and you really CANNOT do much to change it.*	4.99	1.34	1 - 6	-1.31	0.78
(ib2) Your intelligence is something about you that you CANNOT change very much.*	4.99	1.25	1 – 6	-1.21	0.67
(effort_a) If I have to work hard in school, it makes me feel bad about myself.*	5.08	1.11	1 – 6	-1.70	3.11
(effort_b) I often work hard in school.	4.97	0.97	1 – 6	-1.47	3.28
(effort_c) I know I can learn in school if I put in the effort.	4.89	1.28	1 – 6	-1.16	0.58
(persist_a) I get frustrated if something takes a long time to learn*	3.31	1.55	1 – 6	0.22	-1.02
(persist_b) I give up when things take a long time to learn*	4.49	1.40	1 – 6	-0.77	-0.40
(persist_c) When things are hard to learn, it is important to keep trying until you learn it.	4.96	1.09	1 – 6	-1.60	3.05
(mist_a) I am afraid of making mistakes in school.*	3.78	1.61	1 – 6	-0.21	-1.21
(mist_b) I learn a lot from my mistakes or the mistakes of others.	5.04	0.93	1 – 6	-1.76	4.94
(mist_c) Making mistakes is an opportunity for learning.	4.99	1.03	1 – 6	-1.41	2.78
(chall_a) I like it when school is challenging.	4.05	1.56	1 – 6	-0.53	-0.75
(chall_b) I often challenge myself in school.	4.14	1.28	1 – 6	-0.81	0.11
(chall_c) I like challenging material more than easy material.	4.07	1.42	1 – 6	-0.52	-0.48
(learn_a) I get frustrated because I don't know what the best ways are to learn in school.*	4.25	1.47	1 – 6	-0.63	-0.70
(learn_b) When I get stuck, I try to figure it out before I ask for help.	4.78	1.22	1 – 6	-1.30	1.29
(learn_c) I'm not sure how to best learn in school.*	4.02	1.43	1 - 6	-0.34	-0.85

Note. Parenthetical abbreviations will be used in subsequent tables. ib = intelligence belief (Dweck, 1999), effort = effort dimension item, persist = persistence dimension item, mist = mistake dimension item, chall = challenge dimension item, learn = learning dimension item; a = affect, b = behavior, c = cognition. * indicates a reverse scored item.

To evaluate the normality of the items, skewness and kurtosis values were collected with the descriptive statistics in Table 4. Following the suggestions of West, Finch, and Curran (1995) normality was assessed using the benchmarks of |2.0| for skewness and |7.0| for kurtosis. None of the items exceeded either benchmark indicating the items displayed approximately normal distributions.

Item Correlations. Inter-item correlations also were estimated (see Table 5). The two Intelligence belief items were strongly correlated with each other (r = .72), as were the three challenge items (r = .63 – .76). Items on the other theorized dimensions showed weak to moderate inter-item correlations (r = .16 – .53). In addition, items also displayed weak to moderate correlations across dimensions (r = .02 – .57).

Table 5

Inter-item correlations for the MGM measure.

Item	*M* (*SD*)	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17
Q1 ib1	4.99 (1.34)	-																
Q2 ib2	4.99 (1.25)	0.72	-															
Q3 effort_a	5.08 (1.11)	0.30	0.30	-														
Q4 effort_b	4.97 (0.97)	0.23	0.2	0.27	-													
Q5 effort_c	4.89 (1.28)	0.46	0.33	0.53	0.27	-												
Q6 persist_a	3.31 (1.55)	0.26	0.28	0.23	0.16	0.29	-											
Q7 persist_b	4.49 (1.40)	0.26	0.35	0.3	0.28	0.44	0.47	-										
Q8 persist_c	4.96 (1.09)	0.26	0.15	0.17	0.57	0.36	0.17	0.33	-									
Q9 mist_a	3.78 (1.61)	0.18	0.21	0.35	0.02	0.37	0.37	0.43	0.15	-								
Q10 mist_b	5.04 (0.93)	0.15	0.23	0.27	0.42	0.32	0.27	0.39	0.53	0.26	-							
Q11 mist_c	4.99 (1.03)	0.26	0.32	0.31	0.29	0.35	0.18	0.24	0.38	0.3	0.52	-						
Q12 chall_a	4.05 (1.56)	0.3	0.3	0.3	0.4	0.38	0.33	0.4	0.44	0.16	0.31	0.29	-					
Q13 chall_b	4.14 (1.28)	0.24	0.28	0.26	0.4	0.32	0.24	0.25	0.42	0.13	0.22	0.25	0.63	-				
Q14 chall_c	4.07 (1.42)	0.34	0.38	0.33	0.38	0.4	0.31	0.44	0.42	0.15	0.32	0.31	0.76	0.64	-			
Q15 learn_a	4.25 (1.47)	0.39	0.35	0.35	0.17	0.52	0.57	0.51	0.3	0.42	0.36	0.3	0.44	0.31	0.4	-		
Q16 learn_b	4.78 (1.22)	0.2	0.08	0.22	0.33	0.31	0.2	0.24	0.36	0.2	0.29	0.21	0.25	0.29	0.17	0.31	-	
Q17 learn_c	4.02 (1.43)	0.3	0.29	0.3	0.14	0.4	0.31	0.32	0.22	0.27	0.21	0.18	0.2	0.2	0.23	0.4	0.22	-

Note. A heat map was created to help with ease of interpretation: dark green = 1.0, white = 0.0. Underlined correlation coefficients represent inter-correlations within theorized subscales. Bolded correlation coefficients represent high collinearity between items (>.79).

Exploratory Factor Analysis Overview. Before conducting the EFA, it is important to identify whether there are strong enough relationships among items to justify conducting a factor analysis procedure. If the relationships among the items are not strong enough, it will not be possible to obtain a set of factors for the items (Pett, Lackey, & Sullivan, 2003). This threat harkens back to the purpose of a factor analysis. EFA is designed to answer the following question: To what extent do certain constructs (dimensions) explain the observed pattern of correlations? Or, how many constructs are present based on the observed set of responses? The goal of EFA is to reduce items into a smaller set of broader, more generalizable constructs (McCoach, Gable, & Madura, 2013). To the extent that items are unrelated, more factors would be needed to explain the pattern of correlations. Thus the purpose of conducting EFA is to determine the most parsimonious, or simple, interpretable set of factors for a given sample.

In order to determine if the observed correlation matrix is adequate for the factor analysis procedure, there are two common tests that can be conducted: Bartlett's test of sphericity and the Kaiser-Meyer-Olkin test (KMO). Bartlett's test of sphericity tests the hypothesis that the observed correlation matrix is an identity matrix. An identity matrix is one in which the diagonal values are all one and the off-diagonal values are all zero. In other words, the items are all completely unrelated. In this context a researcher would want to reject the null hypothesis to find significant evidence that the observed correlation matrix is not an identity matrix. The KMO test compares the calculated correlation coefficients to partial correlation coefficients. Partial correlations represent the strength of the relationship between two items after removing the effects of all other items. If items load onto a common factor, then removing the effects of those other items would result in a small partial correlation. The KMO test compares the calculated correlations to partial correlations and produces a value between zero and one. Values closer to one indicate stronger relationships between items, while values closer to zero indicate weaker relationships between items. Kaiser (1974) proposed criteria for the KMO statistic that .90 and higher is “marvelous,” .89 to .80 is “meritorious,” .79 to .70 is “middling,” .69 and lower is “unacceptable” (pg. 35).

After the researcher determines that the data are adequate for factoring, the next step is to identify the factor extraction method. There are two common methods of extraction: principal component analysis and common factor analysis (Pett et al., 2003). According to Bandalos (2018), the goal of principal component analysis is to reduce the observed variables to a smaller number of components that contain as much of the variables’ variance as possible. The goal of common factor analysis is to identify latent constructs that explain the relationships (correlations) between observed variables. The goal of this study in the structural stage is to determine whether the latent constructs underlying the observed variables align with theorized dimensions. Therefore, I am using common factor analysis. Specifically, I will be using the most common form of common factor analysis, principal axis factoring.

Next, the researcher needs to determine how many factors to retain. There are several methods that a researcher can use to make this judgement (see Bandalos, 2018; McCoach et al, 2013; and Pett et al., 2003).

The first and perhaps most commonly used criteria is eigenvalue greater than one, also known as the “K1” criterion or Kaiser-Guttman rule. Eigenvalues represent the amount of variance amongst all the items that can be explained by a given factor. An eigenvalue greater than one indicates that the factor explains more than an item's worth of variance.

A second criterion is the percent of variance explained by the factor. It is calculated directly from eigenvalue by dividing the eigenvalue by the number of variables and multiplying the result by 100. This criterion is commonly evaluated through the cumulative percentage or the percent explained by each factor individually (Pett et al., 2003). This method is problematic because of its influence from the number of items and because the cutoff points assigned may be arbitrary, so I will not be using it in this analysis.

A third common criterion utilizes the scree plot. The scree plot is a graph in which the eigenvalue is plotted on the Y axis and the factor number on the X axis. The first factor extracted will always have the largest eigenvalue, with each subsequent factor's eigenvalue decreasing. Typically, there is a point, often referred to as the elbow, where the eigenvalues level out. The common rule is to count the number of factors before the elbow. The next two criteria are statistically based.

The fourth criterion is a test called parallel analysis. Similar to the K1 criterion, parallel analysis relies on eigenvalues. However, instead of comparing eigenvalues to 1, they are compared to sets of randomly produced data. There are two common rules to evaluate the parallel analysis test. The first is comparing observed eigenvalues to the mean of randomly produced eigenvalues. The number of factors retained is determined by the number of factors with an eigenvalue greater than the mean of the random data. The second rule is using the 95^{th} percentile of the random data as a comparison resulting in a more conservative test (McCoach et al., 2013). I will use the 95th percentile as the comparison in the present study.

Another test used to determine the number of factors is the minimum average partial test (Velicer, 1976). Basically, this test works by conducting a factor analysis. After each factor is extracted, the average squared partial correlation between each pair of items is obtained. The number of factors that produces the smallest average squared partial correlation is considered the optimal solution. A revised test (Velicer et al., 2000), which utilizes the 4th power instead of squaring the partial correlation, has been shown to be more accurate, so I will use the revised test in the present study. In summary, the present study will utilize four criteria in determining the appropriate number of factors to extract: (1) K1 criterion, (2) scree plot, (3) parallel analysis test, and (4) minimum average partial test.

Following extraction, the next step in factor analysis is rotation. There are a number of rotation methods, however, all of them fall into two categories: orthogonal and oblique (Bandalos, 2018; Pett et al., 2003). The major difference between the two categories is that orthogonal assumes the factors are uncorrelated, while oblique allows factors to be correlated (they do not have to be correlated). While I am expecting the items on the MGM to fall into six dimensions, because all of these represent different aspects of the same general construct, growth mindset, I am also expecting the dimensions to be correlated. Therefore, I will be utilizing an oblique method of factor rotation.

The final step of the factor analysis is interpretation. When conducting an oblique rotation, factor loadings are presented in pattern and structure matrices. Through partial correlations, the pattern matrix represents the relationship of each item to a factor after removing the effects of other factors. The structure matrix is composed of zero-order, bivariate correlations for each item to each factor. Bandalos (2018) suggested reporting and considering both the pattern and structure matrices in interpretations. When interpreting factor loadings, a common rule is that primary loadings should be above .3 (Bandalos, 2018, Pett et al., 2003). Factor loadings operate similarly to correlations in that squaring them results in the amount of variance explained. Thus, a value over .3 indicates about 10% variance explained. I will be following the suggestion from McCoach et al. (2013) whereby primary loadings will be evaluated against .4 and cross loadings will be evaluated against .3.

Exploratory Factor Analysis Results. An exploratory factor analysis (EFA) was conducted because the proposed items on the MGM measure were new and there was no prior evidence for how these items and their hypothesized dimensions would operate together. Besides the two growth mindset items measuring the intelligence belief dimension, five subscales are being piloted: effort, persistence, challenge, mistakes, and learning strategies. Conducting the EFA allowed me to discover the factor structure present in the current sample and compare that to the theorized dimensions.

Before conducting the EFA procedure, I first determined whether the observed correlation matrix from the present sample of responses was adequate for factoring. To determine this, I conducted both Bartlett's test of sphericity and the KMO test. Bartlett's test of sphericity was significant ($\chi^2 = 1065$, $df = 120$, $p < .001$) indicating there was enough evidence to conclude that the observed correlation matrix was not an identity matrix. The KMO test resulted in a value of .85. According to the suggested criteria from

Kaiser (1974), the value indicated that the correlation matrix was meritorious for a factor analysis procedure.

As indicated in the EFA overview, the present exploratory factor analysis utilized the principal axis factor extraction procedures and employed the direct oblimin rotation method. After determining the procedures for the factor analysis, the next step was to determine the appropriate number of factors to be extracted. I compared the number of factors suggested by four different methods: (1) eigenvalues >1, (2) scree plot – above elbow, (3) parallel analysis, and (4) minimum average partial test. I gave more weight to the parallel analysis and minimum average partial tests because they have been shown to be more accurate (Zwick & Velicer, 1986). Table 7 provides a summary of the number of factors suggested by each method. For the results of each method see Appendix C. The results showed that eigenvalue, scree plot, and the parallel analysis tests suggested a four factor solution. The minimum average partial test suggested a six factor solution, differing from the parallel analysis test.

Table 6

Summary of the number of factors suggested by each method.

Method	Number of Factors
Eigenvalues > 1	4
Scree Plot, above elbow	4
Parallel Analysis, 95th percentile	4
Minimum Average Partial Procedure, 4th power	6

Following the suggestion of Bandalos (2018), when the methods to determine the appropriate number of factors differ, it is recommended to compare results of each suggested factor solution to determine which is more appropriate. Thus, I extracted four and six factor solutions using principal axis factoring and direct oblimin rotation.

Four Factor Results. The results of the four factor solution are presented in Tables 8, 9, and 10. The pattern matrix (Table 9) indicated that the first factor contained seven items. Two of the items had pattern coefficients below.4 on their primary factor. Reviewing the items that loaded onto this factor did not reveal a clear pattern that might suggest why the items are loading onto that factor. Overall, the factor did not make theoretical sense or practical sense. The second and third factors made more theoretical sense. The second factor was composed of the three challenge items, and the third factor was composed of the two intelligence belief items. The fourth factor contained five items. One item displayed a pattern coefficient below .4. The items loading onto this factor did not make theoretical sense but reviewing the items more closely suggested a new possible factor: continual effort. The reason I did not name this factor persistence was to help distinguish it because it contained only one persistence item. In sum, the pattern matrix (Table 9) indicated that most items loaded strongly onto their primary factor (pattern coefficients > .4) with no cross loadings present (> .3). However, the structure matrix (Table 10) indicated that all items had at least one moderate relationship with a second factor and many of them had two or three. Altogether, these results indicated that the four factor solution was not closely aligned to the theorized factor structure, and while the pattern coefficients indicated clean factor loadings, the structure coefficients tempered conclusions about the overall factor structure.

Table 7

Total variance explained by the four factor solution.

	Initial Eigenvalue			Extracted Sums of Squared Loadings		
Factor	Total	% Variance	Cumulative %	Total	% Variance	Cumulative %
I	6.11	35.9	35.9	5.65	33.2	33.2
II (Challenge)	1.74	10.2	46.2	1.29	7.60	40.8
III (Intelligence Belief)	1.35	7.95	54.1	.958	5.64	46.5
IV (Continual Effort)	1.16	6.84	60.9	.764	4.50	51.0

Table 8

Rotated factor pattern matrix to the four factor solution.

	Factor			
Item	I	II	III	IV
(learn_a) I get frustrated because I don't know what the best ways are to learn in school.*	**.672**	-.153	-.077	.009
(mist_a) I am afraid of making mistakes in school.*	**.666**	.119	.025	.046
(persist_a) I get frustrated if something takes a long time to learn.*	**.605**	-.164	.011	-.077
(persist_b) I give up when things take a long time to learn.*	**.558**	-.162	.008	.113
(effort_c) I know I can learn in school if I put in the effort.	**.400**	-.059	-.244	.197
(learn_c) I'm not sure how to best learn in school.*	**.377**	.001	-.169	.065
(effort_a) If I have to work hard in school, it makes me feel bad about myself.*	**.331**	-.058	-.176	.136
(chall_a) I enjoy it when school is challenging.	.114	**-.814**	.018	.029
(chall_c) I like challenging material more than easy material.	.091	**-.801**	-.084	-.006
(chall_b) I often challenge myself in school.	-.013	**-.675**	-.026	.110
(ib1) You have a certain amount of intelligence and you really can't do much to change it.*	-.053	.025	**-.945**	.007
(ib2) Your intelligence is something about you that you can't change very much.*	.059	-.062	**-.741**	-.051
(persist_c) When things are hard to learn, it is important to keep trying until you learn it.	-.103	-.153	.017	**.749**
(mist_b) I am always willing to try and learn from my mistakes in class.	.179	.092	.055	**.701**
(effort_b) I often work hard in school.	-.207	-.203	-.049	**.642**
(mist_c) Making mistakes is an opportunity for learning.	.117	.053	-.147	**.472**
(learn_b) When I get stuck, I try to figure it out before I ask for help.	.158	-.027	.023	**.378**

Note. Obtained using principal axis factoring and direct oblimin rotation. Explains 48% of variance. Bold values indicate primary factor loading. Underlined values indicate strong cross loadings >.3

Table 9

Rotated factor structure matrix to the four factor solution.

Item	Factor I	II	III	IV
(learn_a) I get frustrated because I don't know what the best ways are to learn in school.*	**.758**	-.400	-.433	.396
(mist_a) I am afraid of making mistakes in school.*	**.654**	-.392	-.337	.427
(persist_a) I get frustrated if something takes a long time to learn.*	**.637**	-.107	-.240	.263
(persist_b) I give up when things take a long time to learn.*	**.620**	-.315	-.291	.257
(effort_c) I know I can learn in school if I put in the effort.	**.610**	-.374	-.510	.480
(learn_c) I'm not sure how to best learn in school.*	**.485**	-.296	-.390	.365
(effort_a) If I have to work hard in school, it makes me feel bad about myself.*	**.479**	-.214	-.357	.283
(chall_a) I enjoy it when school is challenging.	.378	**-.859**	-.347	.470
(chall_c) I like challenging material more than easy material.	.381	**-.858**	-.422	.454
(chall_b) I often challenge myself in school.	.261	**-.734**	-.310	.444
(ib1) You have a certain amount of intelligence and you really can't do much to change it.*	.358	-.315	**-.915**	.296
(ib2) Your intelligence is something about you that you can't change very much.*	.384	-.333	**-.773**	.259
(persist_c) When things are hard to learn, it is important to keep trying until you learn it.	.257	-.480	-.251	**.774**
(mist_b) I am always willing to try and learn from my mistakes in class.	.424	-.287	-.229	**.713**
(effort_b) I often work hard in school.	.153	-.469	-.253	**.670**
(mist_c) Making mistakes is an opportunity for learning.	.366	-.271	-.341	**.546**
(learn_b) When I get stuck, I try to figure it out before I ask for help.	.318	-.254	-.186	**.451**

Note. Obtained using principal axis factoring and direct oblimin rotation. Explains 48% of variance. Bold values indicate primary factor loading. Underlined values indicate cross loadings >.3

Six Factor Results. A second EFA was conducted extracting the six factor solution suggested by the minimum average partial test. The results are presented in Tables 11, 12, and 13. The pattern matrix (Table 12) indicated that the first factor contained four items. Again this factor did not align with the theorized factor structure, however it did make practical sense. Reviewing the items revealed that all of them were negatively oriented: “I get frustrated,” “I give up,” or “I am afraid.” In addition, it should be noted that three of the items were affect items and one was a behavior item. The content of the items suggested that they were all getting at negative emotion. Even the behavior item, could be considered to have an affective component. Specifically, giving up could be interpreted as a negative emotional response. Taking into account that the negatively oriented items were reverse scored, I called this factor lack of negative emotion. Another item that seemed most strongly related to the first factor was the cognitive learning strategies item (*I’m not sure how to best learn in school. *Reversed*). However, this item did not display a pattern coefficient above .4 for any factor indicating that it did not have any strong primary loadings. Because the learning strategies cognitive item did not show any primary loadings, it was was excluded from future analyses. In contrast, the second through fifth factors all aligned well with theory. Notably factor two contained all three challenge items as it had in the four factor solution. The third, fourth, and fifth factors contained two intelligence belief, mistake, and effort items respectively. The sixth factor was composed of three items. While the factor did not align with the theorized factor structure, reviewing the three items revealed that the items all related to continual effort, similarly to the four factor results. To differentiate the fifth factor and sixth factor more, because both asked about effort, I reviewed the items loading on the fifth factor. As noted earlier, both were from the effort dimension, and both seemed most focused on the importance of effort so I named the fifth factor effort importance. The structure matrix displayed similar results to the four factor solution where most items displayed a moderate relationship to at least one secondary factor. Further, many of the items displayed moderate relationships to multiple secondary factors. Overall, these results indicated that the factor structure for the six factor solution was much more closely aligned to the theorized structure than the four factor solution. While the structure coefficients temper conclusions for the observed factor structure, I will continue with the six factor structure to explore additional analyses below. Descriptive statistics, factor correlations, and internal consistency are presented for the six factor solution in Table 14.

Table 10

Total variance explained by the six factor solution.

	Initial Eigenvalue			Extracted Sums of Squared Loadings		
Factor	Total	% Variance	Cumulative %	Total	% Variance	Cumulative %
I (Lack of Negative Emotion)	6.11	35.93	35.93	5.71	33.61	33.61
II (Challenge)	1.74	10.22	46.15	1.33	7.83	41.43
III (Intelligence Belief)	1.35	7.95	54.10	1.02	6.00	47.43
IV (Mistakes)	1.16	6.84	60.93	.814	4.79	52.22
V (Effort)	0.95	5.57	71.72	.482	2.84	55.06
VI (Continual Effort)	0.89	5.22	75.93	.401	2.36	57.42

Table 11

Rotated factor pattern matrix to the six factor solution.

	Factor					
Item	I	II	III	IV	V	VI
(lean_a) I get frustrated because I don't know what the best ways are to learn in school.*	**.730**	-.090	-.051	-.017	-.114	.006
(persist_a)I get frustrated if something takes a long time to learn.*	**.614**	-.099	-.098	-.006	.124	.085
(persist_b) I give up when things take a long time to learn.*	**.524**	-.146	-.032	.138	.040	.042
(mist_a) I am afraid of making mistakes in school.*	**.431**	.089	.054	.168	<u>.308</u>	-.097
(learn_c) I'm not sure how to best learn in school.*	.270	.025	-.151	-.046	.220	.137
(chall_c) I like challenging material more than easy material.	.056	**-.879**	-.045	.049	.038	-.124
(chall_a) I enjoy it when school is challenging.	.118	**-.813**	.027	-.013	.015	.021
(chall_b) I often challenge myself in school.	-.018	**-.686**	-.005	-.065	.062	.140
(ib1) You have a certain amount of intelligence and you really CANNOT do much to change it.*	.002	.048	**-.857**	-.103	.076	.146
(ib2) Your intelligence is something about you that you CANNOT change very much.*	.060	-.062	**-.838**	.152	-.056	-.194
(mist_b) I am always willing to try and learn from my mistakes in class.	.162	.042	.027	**.731**	-.053	.191
(mist_c)Making mistakes is an opportunity for learning.	-.049	-.048	-.113	**.560**	.173	-.009
(effort_a) If I have to work hard in school, it makes me feel bad about myself.*	-.031	-.140	-.042	.090	**.606**	-.027
(effort_c) I know I can learn in school if I put in the effort.	.087	-.088	-.133	-.008	**.593**	.178
(Persist_c) When things are hard to learn, it is important to keep trying until you learn it.	-.002	-.166	-.030	.294	-.082	**.580**
(effort_b) I often work hard in school.	-.125	-.230	-.071	.218	-.036	**.482**
(learn_b) When I get stuck, I try to figure it out before I ask for help.	.133	.007	.021	.014	.137	**.444**

Note. Obtained using principal axis factoring and direct oblimin rotation. Explains 48% of variance. Bold values indicate primary factor loading. Underlined values indicate strong cross loadings >.3

Table 12

Rotated factor structure matrix of the six factor solution.

Item	Factor I	II	III	IV	V	VI
(lean_a) I get frustrated because I don't know what the best ways are to learn in school.*	**.754**	-.395	-.402	.302	.499	.297
(persist_a)I get frustrated if something takes a long time to learn.*	**.717**	-.307	-.285	.214	.275	.172
(persist_b) I give up when things take a long time to learn.*	**.650**	-.400	-.319	.380	.390	.274
(mist_a) I am afraid of making mistakes in school.*	**.563**	-.126	-.195	.308	.499	.075
(learn_c) I'm not sure how to best learn in school.*	.433	-.214	-.327	.174	.417	.244
(chall_c) I like challenging material more than easy material.	.356	**-.890**	-.404	.359	.293	.284
(chall_a) I enjoy it when school is challenging.	.373	**-.848**	-.330	.318	.271	.375
(chall_b) I often challenge myself in school.	.239	**-.732**	-.294	.243	.244	.411
(ib1) You have a certain amount of intelligence and you really CANNOT do much to change it.*	.310	-.323	**-.868**	.161	.387	.261
(ib2) Your intelligence is something about you that you CANNOT change very much.*	.338	-.361	**-.863**	.311	.300	.033
(mist_b) I am always willing to try and learn from my mistakes in class.	.371	-.325	-.203	**.808**	.262	.445
(mist_c)Making mistakes is an opportunity for learning.	.252	-.317	-.314	**.638**	.366	.255
(effort_a) If I have to work hard in school, it makes me feel bad about myself.*	.479	-.389	-.446	.317	**.746**	.390
(effort_c) I know I can learn in school if I put in the effort.	.339	-.323	-.328	.309	**.663**	.203
(Persist_c) When things are hard to learn, it is important to keep trying until you learn it.	.229	-.500	-.237	.537	.190	**.736**
(effort_b) I often work hard in school.	.119	-.485	-.242	.437	.164	**.631**
(learn_b) When I get stuck, I try to figure it out before I ask for help.	.286	-.251	-.154	.240	.298	**.501**

Note. Obtained using principal axis facotring and direct oblimin rotation. Bold values indicate primary factor loading. Underlined values indicate cross loadings >.3

Table 13

Descriptive statistics, factor correlations, and internal consistency for the six factor solution.

Factor	*M*	*SD*	# of items	I	II	III	IV	V	VI	*IC*
I (Lack of Negative Emotion)	3.96	1.16	4	-						.771[a]
II (Challenge)	4.09	1.26	3	-.32	-					.861[a]
III (Intelligence Belief)	4.99	1.20	2	-.34	.38	-				.836[b]
IV (Mistakes)	5.01	0.86	2	.30	-.36	-.24	-			.688[b]
V (Effort Importance)	4.98	1.05	2	.48	-.26	-.37	.29	-		.695[b]
VI (Continual Effort)	4.90	0.86	3	.21	-.41	-.18	.35	.23	-	.673[a]

Note. IC = internal consistency.
[a] Internal consistency calculated using Cronbach's alpha.
[b] Internal consistency calculated using Spearman-Brown

Table 14

Descriptive statistics, factor correlations, and internal consistency for the theorized factor solution.

Factor	*M*	*SD*	# of items	I	II	III	IV	V	VI	*IC*
I (Intelligence Belief)	4.99	1.20	2	-						.836[c]
II (Effort)	4.98	.85	3	.44	-					.629[b]
III (Persistence)	4.25	1.00	3	.38	.54	-				.593[b]
IV (Mistakes)	4.60	.90	3	.32	.51	.57	-			.570[b]
V (Challenge)	4.09	1.26	3	.38	.52	.53	.34	-		.861[b]
VI (Learning Strategy)	4.51	1.09	2[a]	.36	.53	.61	.50	.44	-	.474[c]

Note. Factors are listed in the order items appear on the actual measure. IC = internal consistency.
[a] Learning strategy factor was calculated without the cognitive item present, per the results of the EFA.
[b] Internal consistency calculated using Cronbach's alpha.
[c] Internal consistency calculated using Spearman-Brown.

Internal Consistency Overview. Following the EFAs, I analyzed the internal consistency of both the theorized and observed factors. There are several common methods to examine internal consistency, the most common being Cronbach's alpha. While the alpha coefficient is more common and preferred over other methods, such as split-half reliability, it becomes biased when used with a two-item scale. In this situation, the alpha coefficient will tend to underestimate the true reliability of a scale (Eisinga, Grotenhuis, & Pelzer, 2012). In situations with a two-item scale, the Spearman-Brown coefficient tends to be more accurate (Eisinga et al., 2012). Therefore, I calculated the Cronbach alpha coefficient for factors with more than two items, and I calculated the Spearman-Brown coefficient for factors with two items.

Cronbach's alpha is an indicator of the extent to which items on a scale are measuring the same construct. The alpha coefficient can range from zero to one. Values closer to zero indicate poor internal consistency and values close to one indicate strong internal consistency. Historically, values in the range of .70 - .79 are considered acceptable, .80 - .89 is considered good, and .90 and higher is considered excellent (Nunnally & Bernstein, 1994). However, Nunnally actually went on to say that the context surrounding the measure should also be taken into account. The dimensions of the MGM measure are designed to be short and used for improvement purposes. As the number of items on a scale decreases, the alpha coefficient will also decrease (Cortina, 1993). Recognizing that the alpha coefficient is likely to be artificially smaller, a lower level of reliability may be acceptable (Kosovich, Hulleman, & Barron, 2017; Yeager & Bryk, 2015). Similarly, Spearman-Brown coefficient values also range from zero to one, and we can evaluate them with the same criteria as with Cronbach's alpha. In fact, when used on a two-item scale, the Spearman-Brown coefficient will be equivalent to the standardized alpha coefficient.

Internal Consistency Results. A general comparison of the internal consistency of the theorized dimensions versus observed factor solution revealed that the observed factor solution displayed higher internal consistency. This is not surprising, given that the purpose of the factor analysis is to explain the pattern of relationships present in the sample. For the theorized dimensions, only the challenge (α = .86) and intelligence belief (ρ = .84) factors displayed acceptable levels of reliability. It should be noted that these factors were identical to the observed factor structure, thus their reliability values were the same. The effort (α = .63), persistence (α = .59), mistakes (α = .57), and learning Strategy (ρ = .47) dimensions all displayed reliability below .70. As previously indicated, lower reliability values may be accepted, however values in the .50 range and lower may be too low.

For the observed factor structure, the challenge factor again displayed the strongest internal consistency with an alpha of .86. This value met acceptable levels of reliability. The intelligence belief (ρ = .84), negative emotion (α = .77), and effort importance (ρ = .70) factors also displayed acceptable levels of reliability. The Mistakes (ρ = .69) and Continual Effort (α = .67) factors did not show a reliability above .70, however, as indicated in the overview, these may be in the acceptable range.

External Stage Analyses

The external stage of Benson's (1998) validity framework is focused on relationships to other constructs. Specifically, Benson described this stage as the most important stage of the process because this is where scores on the measure start to take on meaning. As described by Crocker and Algina (1986), “an operational definition of a construct is not enough; the meaningfulness or importance of the construct must also be made explicit through a description of how it is related to other variables” (p. 230). While there were some reservations from the factor analysis and the internal consistency results, I continued to the external stage following a quote from Cronbach (1961), “If predictive validity is satisfactory, low reliability does not discourage us from using the test” (p. 128).

To gain evidence for the external stage, I analyzed the relationship of the theorized dimensions and the observed factor structures with two education outcomes (Final Grades and SOL scores). The reason for using education outcomes stemmed from the inception of the improvement work being conducted in the local school district. The underlying assumption was that students with stronger growth mindsets would demonstrate increased outcomes (e.g., higher grades and stronger performances on standardized tests). Grades and SOL scores then serve within the family of measures (improvement principle #4 – you cannot improve what you cannot measure) as the ultimate outcomes of interest to determine the efficacy of the improvement project. As highlighted in the Crocker & Algina (1986) and Cronbach (1961) quotes above, the ability of the theorized dimensions or the emergent six-factor structure to predict Grades and SOL scores is perhaps the most important piece of validity evidence to collect.

Additionally, two education outcomes were identified to avoid problems that may interfere with inferences to grades. First, grades often incorporate variables that are not related to scholastic aptitude (e.g., attendance or behavior). Second, because teachers were so closely partnered on this research project, it is conceivable that a teacher’s perception of students' growth mindset could influence their grading. Pearson correlation coefficients between grades English and Math grades and SOL scores were moderate to strong ranging from .52 to .72.

Analyses Overview. First, to examine the general relationship between the theorized and emergent factor structure with English and Math grades and SOL scores I conducted bivariate correlations. Second, to examine how well the theorized dimensions and the emergent factor structure predicted English and Math grades and SOL scores multiple regression models were estimated.

Specifically, regression models for both the theorized dimensions and the observed six factor solution predicting English and Math grades and SOL scores were estimated. Because the focus of the analyses was to determine how well the theorized dimensions and emergent factor structure served as predictors overall, the multiple regression models were estimated via simultaneous entry of the predictors.

Regression models that included item-level predictors were not run to avoid two potential multivariate problems: multivariate power and null washout. Multivariate power is likely to occur when predictors are less related (correlated) to one another and as the number of predictors increases. Null washout occurs when there is at least one significant predictor, but the effect of the predictor is overcome or "washed out" by the inclusion of non-significant predictors. In both multivariate power and null washout, the problem is exacerbated as the number of predictors increases. To reduce the likelihood of multivariate power and null washout while still retaining some degree of useful information, I decided to only use the theorized dimensions and observed factor structure instead of each item individually, meaning the number of predictors is reduced to six instead of 17.

Final Grade Correlations. Descriptive statistics for Final Grades can be found in Table 15. Correlations of theorized dimensions and observed factor structure with final grades are presented in Tables 16 and 17 respectively. Correlations at the item level to final grades are presented in Appendix D. For the theorized dimensions, intelligence belief had the strongest correlation with grades. Following intelligence belief, the challenge dimension showed the next strongest correlation with grades. Mistakes showed the weakest relationship with grades. The same pattern of relationships was found in the observed factor structure as well. Intelligence belief and challenge displayed the strongest correlations with grades and mistakes displayed the weakest correlation with grades.

Table 15

Descriptive statistics for English and Math final grades.

Course	*M*	*SD*	Skewness	Kurtosis
English Final	88.92	7.05	-1.03	1.51
Math Final	91.94	4.29	-0.82	0.81

Note. *N*=149. Final = Final grade in class.

Table 16

Correlations between theorized dimensions and final grades in core classes.

Theorized Dimension	English Final	Math Final
Intelligence Belief	0.41	0.43
Effort	0.35	0.32
Persistence	0.28	0.32
Mistakes	0.15	0.20
Challenge	0.37	0.38
Learning Strategy	0.31	0.29

Note. *N*=149. Final = Final grade in class.

Table 17

Correlations between emergent factors (six factor EFA solution) and final grades in core classes.

Factor	English Final	Math Final
I (Lack of Negative Emotion)	0.23	0.36
II (Challenge)	0.37	0.38
III (Intelligence Belief)	0.41	0.43
IV (Mistakes)	0.17	0.08
V (Effort Importance)	0.26	0.26
VI (Continual Effort)	0.36	0.24

Note. *N*=149. Final = Final grade in class.

Final Grade Multiple Regressions. Two regression models were estimated for the theorized dimensions predicting English and Math final grades. First a simultaneous multiple regression model was estimated predicting final English grades from the six dimensions (see Table 18). The full model, including intelligence belief and total scores from the other dimensions (effort, persistence, mistakes, challenge, and learning strategy), explained 25% of variance in English grade [$R^2 = .251$, $F(6, 142) = 7.936$, $p < .001$]. While the full model was significant, only the intelligence belief ($\beta = .273$, $t = 3.28$, $p = .001$) and challenge ($\beta = .189$, $t = 2.06$, $p = .042$) dimensions were statistically significant predictors of English grades. The squared semi-partials, which indicate the amount of variance in English grade explained uniquely by a predictor, showed that the intelligence belief dimension uniquely explained 5.7% of variance in English grade and challenge explained 2.2%, controlling for the other predictors in the model.

Table 18

Multiple regression of theorized dimensions predicting final English grades.

Dimension	β	t	p	sr	sr^2
Intelligence Belief	.273	3.28	.001	.238	.057

Effort	.135	1.36	.177	.099	.010
Persistence	.007	-.069	.945	.005	.000
Mistakes	-.145	-1.53	.128	-.111	.012
Challenge	.189	2.06	.042	.149	.022
Learning Strategies	.131	1.33	.187	.096	.009

Note. *N*=148; *sr* = semi-partial

Second, a simultaneous multiple regression model was estimated for the six theorized dimensions predicting final Math grades (see Table 19). The full model explained 25% of variance [R^2 = .249, F(6, 142) = 7.857, p < .001]. Again, only the intelligence belief (β = .303, t = 3.63, p < .001) and challenge (β = .206, t = 2.24, p = .027) dimensions were statistically significant predictors of Math grades. The intelligence belief dimension explained 7.0% of variance of Math grade and challenge explained 2.7%, controlling for the other predictors in the model.

Table 19

Multiple regression of theorized dimensions predicting final Math grades.

Dimension	β	t	p	sr	sr^2
Intelligence Belief	.303	3.63	>.001	.264	.070
Effort	.043	.430	.668	.031	.001
Persistence	.077	.720	.473	.052	.003
Mistakes	-.049	-.520	.604	-.038	.001
Challenge	.206	2.24	.027	.163	.027
Learning Strategies	.045	.450	.653	.033	.001

Note. *N*=148; *sr* = semi-partial

Two additional regression models were estimated to analyze the emergent six factor solution predicting both English and Math final grades. First, a simultaneous multiple regression model was estimated to predict English grade from the six emergent factors (see Table 20). The full model, including Negative Emotion, Challenge, Intelligence Belief, Mistakes, Effort Importance, and Continual Effort, explained 27% of variance in English grade [R^2 = .266, F(6, 142) = 8.564, p < .001]. Only the intelligence belief (β = .315, t = 3.787, p < .001) and continual effort (β = .266, t = 2.876, p = .005) factors were statistically significant predictors. Moreover, the intelligence belief dimension uniquely explained 7.4% of variance in English grade and continual effort explained 4.3%, given the other predictors in the model.

Table 20

Multiple regression of emergent factors predicting final English grades.

Factor	β	t	p	sr	sr^2
Lack of Negative Emotion	-.017	-.192	.848	-.014	.000
Challenge	.174	1.932	.055	.139	.019
Intelligence Belief	.315	3.787	< .001	.272	.074
Mistakes	-.117	-1.311	.192	-.094	.009
Effort Importance	.001	.012	.991	.001	.000
Continual Effort	.266	2.876	.005	.207	.043

Note. N=148; sr = semi-partial

Lastly, a simultaneous multiple regression model was estimated with the emergent six factors predicting final Math grade (see Table 21). The full model explained 30% of variance in Math grade [R^2 = .295, F(6, 142) = 9.896, p < .001]. Four factors were statistically significant predictors: Negative Emotion (β = .223, t = 2.505, p = .013), Challenge (β = .220, t = 2.493, p = .014), Intelligence Belief (β = .316, t = 3.880, p < .001), and Mistakes (β = -.221, t = -2.528, p = .013). Intelligence belief was the strongest predictor, explaining 9.6% variance in math grade, uniquely. Negative Emotion, Challenge, and Mistakes explained 4.2%, 4.2%, and 4.3% of variance, respectively. Surprisingly, Mistakes was a significant negative predictor, however it looks like it was a result of suppression when testing these variables simultaneously. In general, suppression occurs when the absolute value of a predictor's standardized regression weight (β) is greater than its bivariate correlation with the criterion (in this case Math Grade) or when the two have different signs (Kline, 2011). Examining the correlation between the

Mistakes factor and Math grade and the β-weight revealed a weak correlation (r=.08) and a significant regression weight (β = -.221), supporting the likelihood of this being a suppression effect. There are also specific types of suppression and it appears that this is a case of classic suppression, where one predictor is uncorrelated with the criterion but displays a non-zero β-weight controlling for another predictor (Kline, 2011).

Table 21

Multiple regression of emergent factors predicting final Math grades.

Factor	β	t	p	sr	sr^2
Lack of Negative Emotion	.223	2.505	.013	.206	.042
Challenge	.220	2.493	.014	.205	.042
Intelligence Belief	.316	3.880	< .001	.310	.096
Mistakes	-.221	-2.528	.013	-.208	.043
Effort Importance	-.042	-.459	.647	-.039	.002
Continual Effort	.100	1.108	.270	.093	.009

Note. N=148; sr = semi-partial

SOL Correlations. Following the analyses of final grades, similar analyses were also conducted for SOL scores. Descriptive statistics for the English and Math SOL tests can be found in Table 22. First, correlations between theorized dimensions and SOL scores were run and are presented in Tables 22 and 23. Correlations at the item level to SOL scores are presented in Appendix E. Generally, it appears that correlations were stronger between theorized dimensions and Math SOL scores than for English SOL scores. Furthermore, the intelligence belief dimension showed the strongest correlation to both English and Math SOL scores and mistakes showed the weakest correlation. Challenge showed the next strongest correlations behind intelligence belief. For the observed factor structure, intelligence belief again showed the strongest correlations and mistakes showed the weakest. Lack of negative emotion and challenge showed the next strongest correlations, especially for the Math SOL.

Table 22

Descriptive statistics for English and Math SOL scores.

SOL	*M*	*SD*	Range	Skewness	Kurtosis
English SOL	437.73	66.27	200 – 600	-0.39	0.36
Math SOL	446.52	58.14	200 – 600	0.42	0.15

Note. *N*=146. Range = theoretical range of possible scores. SOL refers to the standardized tests taken by K-12 students at the end of the school year. In 6th grade the only SOLs that are taken are Reading (English) and Mathematics.

Table 23

Correlations between theorized dimensions and SOL scores.

Dimension	English SOL	Math SOL
Intelligence Belief	0.53	0.53
Effort	0.28	0.41
Persistence	0.29	0.42
Mistakes	0.21	0.30
Challenge	0.28	0.49
Learning Strategies	0.33	0.39

Note. *N* = 146. SOL refers to the standardized tests taken by K-12 students at the end of the school year. In 6th grade the only SOLs that are taken are Reading (English) and Mathematics.

Table 24

Correlations between observed six factor solution and SOL scores.

Factor	English SOL	Math SOL
Lack of Negative Emotion	0.35	0.44
Challenge	0.28	0.49
Intelligence Belief	0.53	0.53
Mistakes	0.16	0.25
Effort Importance	0.27	0.35
Continual Effort	0.18	0.34

Note. *N* = 146. SOL refers to the standardized tests taken by K-12 students at the end of the school year. In 6th grade the only SOLs that are taken are Reading (English) and Mathematics.

SOL Multiple Regressions. Multiple regression models were estimated for the six theorized dimensions predicting English and Math SOL scores. First, a multiple regression model was estimated with the six theorized dimensions predicting English

SOL scores (see Table 24). The full model explained 31% of variance [$R^2 = .308$, $F(6, 139) = 10.314$, $p < .001$]. Only the intelligence belief dimension ($\beta = .472$, $t = 5.830$, $p < .001$) was a statistically significant predictor. The squared semi-partial indicated that intelligence belief uniquely explained 16.9% of variance in English SOL score, given the other predictors in the model.

Table 25

Multiple regression of theorized dimensions predicting English SOL scores.

Dimension	β	*t*	*p*	*sr*	sr^2
Intelligence Belief	.472	5.830	<.001	.411	.169
Effort	-.038	-.397	.692	-.028	.001
Persistence	.032	.310	.757	.022	.000
Mistakes	-.038	-.418	.677	-.029	.001
Challenge	.057	.635	.526	.045	.002
Learning Strategies	.155	1.628	.106	.115	.013

Note. *N*=148; *sr* = semi-partial. SOL refers to the standardized tests taken by K-12 students at the end of the school year. In 6th grade the only SOLs that are taken are Reading (English) and Mathematics.

Second, a multiple regression model was estimated with the six theorized dimensions predicting Math SOL scores (see Table 25). The full model explained 40% of variance [$R^2 = .399$, $F(6, 139) = 15.364$, $p < .001$]. The intelligence belief ($\beta = .365$, $t = 4.842$, $p < .001$) and challenge ($\beta = .259$, $t = 3.115$, $p = .002$) dimensions were found to be statistically significant predictors. The intelligence belief dimension explained 10.1% of variance of Math SOL score and challenge explained 4.2%, given the other predictors in the model.

Table 26

Multiple regression of theorized dimensions predicting final Math SOL scores.

Dimension	β	*t*	*P*	*sr*	sr^2
Intelligence Belief	.365	4.842	< .001	.318	.101
Effort	.031	.345	.730	.023	.001

Persistence	.072	.738	.462	.049	.002
Mistakes	-.005	-.064	.949	-.004	.000
Challenge	.259	3.115	.002	.205	.042
Learning Strategies	.094	1.062	.290	.070	.005

Note. *N*=148; *sr*=semi-partial. SOL refers to the standardized tests taken by K-12 students at the end of the school year. In 6th grade the only SOLs that are taken are Reading (English) and Mathematics.

Next, two multiple regression models were estimated for the emergent six factor structure predicting English and Math SOL scores. First a regression model was estimated with the six factors predicting English SOL scores (see Table 26). The full model explained 32% of variance [$R^2 = .318$, $F(6, 139) = 10.803$, $p < .001$]. Both the Negative Emotions ($\beta = .203$, $t = 2.250$, $p = .026$) and intelligence belief factors ($\beta = .471$, $t = 5.816$, $p < .001$) were statistically significant predictors. Intelligence belief uniquely explained 16.6% of variance in English SOL score and Negative Emotions explained 2.5%, given the other predictors in the model.

Table 27

Multiple regression of emergent factors predicting English SOL scores.

Dimension	β	t	p	sr	sr^2
Lack of Negative Emotion	.203	2.250	.026	.158	.025
Challenge	.086	.972	.332	.068	.005
Intelligence Belief	.471	5.816	< .001	.407	.166
Mistakes	-.105	-1.225	.223	-.086	.007
Effort Importance	-.025	-.276	.783	-.019	.000
Continual Effort	-.024	-.268	.789	-.019	.000

Note. *N*=148; *sr*=semi-partial. SOL refers to the standardized tests taken by K-12 students at the end of the school year. In 6th grade the only SOLs that are taken are Reading (English) and Mathematics.

Lastly, a multiple regression model was estimated with the six factors predicting Math SOL scores (see Table 27). The full model explained 41% of variance [$R^2 = .413$, $F(6, 139) = 15.364$, $p < .001$]. The Negative Emotion ($\beta = .212$, $t = 2.532$, $p = .012$), Challenge ($\beta = .279$, $t = 3.392$, $p = .001$), and Intelligence Belief ($\beta = .375$, $t = 4.992$, $p <$

.001) factors were found to be statistically significant predictors. The Intelligence Belief factor was the strongest predictor, uniquely explaining 10.5% of variance of Math SOL score, given the other predictors in the model. Challenge and Negative Emotion explained 4.8% and 2.7% of variance, respectively.

Table 28

Multiple regression of emergent factors predicting Math SOL scores.

Dimension	β	t	p	Sr	sr^2
Negative Emotion	.212	2.532	.012	.165	.027
Challenge	.279	3.392	.001	.220	.048
Intelligence Belief	.375	4.992	.000	.324	.105
Mistakes	-.090	-1.133	.259	-.074	.005
Effort Importance	-.027	-.318	.751	-.021	.000
Continual Effort	.047	.571	.569	.037	.001

Note. N=148; sr=semi-partial SOL refers to the standardized tests taken by K-12 students at the end of the school year. In 6^{th} grade the only SOLs that are taken are Reading (English) and Mathematics.

Discussion

Following Benson's (1998) three stages of a strong program of construct validation, the purpose of this study was to develop a new, multi-dimensional measure of growth mindset (Dweck, 2006). To complete the first stage, *substantive stage,* I followed the steps outlined in Gehlbach and Brinkworth (2011) that resulted in a measure theorized to measure six dimensions of growth mindset: intelligence belief, effort, persistence, mistakes, challenge, and learning strategy. After creating the measure, I began the investigation of the structural and external stages to gain psychometric evidence for the MGM (Multi-dimensional Growth Mindset) measure.

Structural Stage

The results of the six factor solution showed moderate support for the theorized dimensions. Factors 2, 3, 4, and 5 included items cleanly aligning with 4 of the 6 proposed dimensions. Specifically, factor 2 included all three challenge items, factor 3 included the two intelligence belief items, factor 4 included two out of the three mistake items and factor 5 included two of the three items from the effort dimension. Two of the theorized factors, persistence and learning strategies, did not have items loading cleanly onto their own dimensions. Instead, factors 1 and 6 included items from various dimensions. After analyzing the content of the items for factors 1 and 6 more closely, an alternative hypothesis emerged for why these items may be holding together in specific ways. Factor 1 included only negatively oriented items about emotional (e.g., *I get frustrated if something takes a long time to learn.*Reversed*). Taking into account reverse scoring, I named this factor Lack of Negative Emotion. Factor 6 included items that appeared to relate to continual effort over time. Specifically this factor included one persistence item (*When things are hard to learn, it is important to keep trying until you learn it.*), one effort item (*I often work hard in school.*), and one learning strategy item (*When I get stuck, I try to figure it out before I ask for help.*). While this may seem like it would be covered under the purview of persistence, I did not name it persistence because only one item on the factor was a persistence item. Thus, I wanted to differentiate this factor from Persistence and named it Continual Effort.

In terms of the theorized dimensions, if the measure was operating as intended, there would be five dimensions with three items each and the intelligence belief dimension with two items. The three items within the five theorized dimensions would represent students' affect, behavior, and cognition (Dweck & Leggett, 1988) toward a dimension. The major question to ask then, is what was the emergent structure of the items? As previously discussed, several affect items broke out to form the lack of negative emotion factor. There are two possible reasons for this. The first is that the items were negatively oriented, and the second is that affect items operated differently than behavior or cognition items.

First, affect items may not have operated as intended because they were negatively oriented. Bandalos (2018) indicated that including both positive and negative oriented items may change the dimensionality of the scale producing “method effects.” Essentially, rather than items grouping into theorized dimensions, they group into positive vs. negative categories of items. For example, Factor 1 of the present study contains only negatively oriented items. In addition to items operating differently, research has shown additional challenges to negatively oriented items. One challenge is that they can be confusing for test takers (Sherman, 1973; Wason, 1959). Another is that ability to respond to negatively oriented items may be related to reading levels (Marsh, 1986). This last challenge is particularly noteworthy for the current study because of the unique demographic makeup of the Harrisonburg City Public School system. Many students come from other countries or come from homes where a language other than English is the primary language. Approximately 16% of the students in the present sample were identified as Limited English Proficiency. In addition, I needed to provide read aloud support to two of the nine math classes (read each item out loud with all students in the class following along). Thus, the negatively oriented items may have proven particularly confusing for this population of students.

Second, the affect items may not have operated as intended because they function differently from behavior and cognition items. However, because the affect items were confounded with negative orientation (four out of the five were negatively oriented), there is no way to tease apart whether the unusual relationships were due to the negative wording or due to something more fundamental with respect to affect.

In addition to the affect items, the results from the structural stage also indicated other challenges to the MGM measure. One such challenge was writing items that only assessed one dimension. This is highlighted in the inter-item correlation matrix (Table 5). As previously noted, only the intelligence belief and challenge scores showed strong inter-dimensional correlations, which was supported by the factor analysis. On the other hand, scores from the other theorized dimensions (Effort, Persistence, Mistakes, and Learning Strategies) showed moderate inter-dimensional correlations that were of similar magnitude to correlations with scores from other dimensions. For example, the effort dimension correlations ranged from .27 to .53. Correlations between the effort item scores with scores from other dimensions ranged from .02 to .57 with most of the correlations falling between .20 and .40. Scores from two items from the mistakes and effort dimensions formed their own respective factors, but the rest of the items were spread among factors 1 and 6 or didn't load onto any factor in the case of the learning strategy item (*I'm not sure how to best learn in school*.* *Reversed*). Together these results may indicate that the dimensions are not disparate enough, as defined, for survey items to adequately measure the unique dimensions.

The final concern was with the learning strategy dimension. This dimension had the weakest psychometric support. Scores from two learning strategy items loaded separately onto Factors 1 and 6, which are the factors formed from items of various dimensions. The third learning strategy item did not have a strong primary loading on any factor.

We gained some insight into this problematic dimension during the substantive stage. Recall that I conducted interviews with students. One of the key takeaways was that none of the interviewed students knew what the term learning strategy meant. However, students did recognize common forms of learning strategies (such as studying and flashcards). This forced me to write items that explicitly avoided using the phrase "learning strategy" in the item, making it difficult to understand at face value whether the proposed items measured the learning strategy dimension. For example, scores from the proposed learning strategy item *"When I get stuck, I try to figure it out before I ask for help"* loaded on Factor 6 of the EFA with other items that reflected continual effort rather than learning strategies.

Overall the initial factor analysis did not align with the theorized dimensions for the items as written. Four of the six factors in the solution were composed of items from a single dimension respectively. However, having both positively and negatively oriented items may have introduced a method effect and made the measure more difficult to understand, especially for the population of students used in the current study. Next, the proposed dimensions may be much more related than initially hypothesized. Lastly, the learning strategies items in particular operated poorly, which may have been due to not being able to use the actual phrasing "learning strategy."

External Stage

To build evidence for the external stage of Benson's (1998) validity framework, I estimated multiple regression models predicting grades and SOL scores. Both the theorized dimensions of the MGM and the observed six factor solution to the EFA were used as predictors for English and Math final grades as well as English and Math SOL scores. The purpose of running models with both theorized and observed predictors was to compare their predictive utility and to learn as much as I could from the initial study. As Cronbach (1961) shared "If predictive validity is satisfactory, low reliability does not discourage us from using the test" (p. 128). Therefore, even though the structural stage did not find evidence fully supporting the theorized dimensions, they were used as a comparison against the observed six factors.

The results of the multiple regressions highlighted two overall trends. The first is that both the theorized dimensions and emergent six factors displayed higher R^2 values, explained more variance, for Math grades and SOL scores than for English. This may just A possible explanation may be that the teachers I partnered with in the 6^{th} grade were all math teachers. Most of the growth mindset interventions delivered to students throughout the year occurred in their math classes. Thus, it could be that students developed a stronger growth mindset for Math than for English. Some of the growth mindset interventions were conducted across all 6^{th} grade teachers to increase transferability of the mindset to multiple contexts. However, it should be noted that the measure was completed by students in the context of their math class. Additionally, if the math teachers remained the most devoted to developing students' growth mindsets throughout the year, it is possible that students' mindsets for math were stronger than they were for English.

The second overall trend in the regression results was that the observed six factor solution was consistently a stronger predictor of both grades and SOL scores than the theorized dimensions. The internal consistency results may highlight why. The results of the analysis of internal consistency revealed that the six factor solution demonstrated higher internal consistency than the theorized dimensions. This is not surprising given that the goal of the factor analysis is to group items that share the most variance together, suggesting that they potentially measure the same thing (internal consistency). This is relevant to the multiple regression because in order to conduct the regression I averaged the items to create scores for the observed factors and the theorized dimensions. First, when items are averaged, some of the information held in the items (such as variance) gets lost. Second, when the items were combined in the theorized dimensions, they were not always combined with the items that they shared the most variance with. Thus, when the multiple regression was conducted, some of the unique variance may have been lost by averaging the items with other items that were not optimal to combine with. This in turn led to poorer performing predictors and less overall variance explained for the theorized dimensions.

Taking a closer look at the individual predictor results from the regressions revealed that the intelligence belief and challenge factors were consistently significant predictors of grades and SOL scores. Intelligence belief in every case was the strongest predictor of performance. However, challenge explained additional variance highlighting how it may be important to have both an intelligence belief supported with a spectrum of additional beliefs like a positive view of challenge. This is reflected in an anecdote from Dr. Kenn Barron who always teaches his freshman college students about growth mindset. One year his students commented that they have a growth mindset until things get challenging, further reinforcing that just measuring a belief about intelligence may not be enough to understand when students are optimally motivated.

For the theorized dimensions there were no other dimensions that were significant predictors. On the other hand, the observed six factor solution displayed some additional significant factors. Lack of negative emotion was a significant predictor of final Math grade and both English and Math SOL scores, Continual Effort was a significant predictor of English grade, and Mistakes was a significant negative predictor of Math grade.

The lack of negative emotion pattern of results indicates that individuals who tend to get frustrated or react negatively to the learning environment tend to not do as well in Math class and on the English and Math SOL scores. This makes sense especially in context of a standardized testing session. Doing well on these tests requires a lot of focus and attention, if a student is prone to frustration when something doesn't make sense or when getting an obviously incorrect answer this could certainly lead to a poorer performance.

The continual effort factor was a significant predictor for English, indicating that students who performed well were students who continued to put in effort despite struggling sometimes. This result with this group of students where many are labelled as Limited English Proficiency (LEP) or just have lower reading levels harkens back to a result found in previous research at HCPS (Barron et al., 2017). One of the groups found to have the strongest growth mindset involved students who were former LEP. In other words, students who had been labelled LEP and had since gained strong enough English skills to test out of LEP held strong growth mindsets. Although not tested causally, this suggests that either the students who make it out have a stronger growth mindset to begin with or the process of learning a new language serves as a learning opportunity teaching growth mindset.

The mistake result was noteworthy because it was the only significant *negative* predictor and it occurred for math grade. After examining this finding more deeply I determined it was likely a suppression effect. Comparing the regression result, to the zero-order correlation between Math grade and the mistake factor ($r = .08$) further supported that it may be an instance of classic suppression (Kline, 2011).

In sum, the external stage results displayed three major trends. The first was that the theorized dimensions and emergent six factors explained more variance in Math grades and SOL scores than English, which could have resulted from the math teachers being the most dedicated to teaching growth mindset to students among all of the 6th grade teachers. However, more information would be necessary to support this conclusion. The second major trend was that the emergent six factors were consistently stronger predictors than the theorized dimensions. This may be due to the fact that as a result of the EFA the items were combined in more optimal ways than through the theorized dimensions. The third major trend was that the original intelligence belief items (versus any of the newly proposed dimensions on the MGM measure) continued to be the best predictor of student outcomes.

General Discussion

As with any research project, there are potential implications for theory, research, and practice. I conclude with my thoughts on how the current study can impact all three of these areas.

Implications for Theory

The theoretical impact of this line of research for growth mindset may be quite significant. Recently, growth mindset has come to be regarded by many as a silver bullet in education. Coming off the back of some impressive findings, it seemed that many of the problems faced by educators could be remedied by instilling students with the proper mindset. Recent studies tell a different story, however. A meta-analysis from Sisk et al. (2018) noted that while there are studies showing large positive effects of growth mindset, there are others that show a null or even negative effect of growth mindset. When positive effects were found, it appeared to be mainly students who historically performed more poorly or were academically at risk.

A study from a research lab at the University of Virginia (Wormington et al., under review) may offer another explanation for these findings. As previously discussed, they found that participants who discussed mindset beliefs and learning strategies or additional growth mindset behaviors showed significantly improved performance in classes over the control condition. Carol Dweck herself, even recognized the importance of learning strategies in a recent review article (Dweck & Yeager, 2019) in which she said “some educators told their students that they could do anything but did not provide them with strategies, guidance, or information about resources for the accomplishment of this promise” (pg. 10). Clearly there are other dimensions that are important for growth mindset. Dweck and others are also now suggesting that holding a belief in your intelligence with no understanding of how to act on that mindset is a “false growth mindset.”

The current study supports this disposition. A group of math teachers noticed that their students were not demonstrating growth mindset behaviors following a series of growth mindset interventions. Looking at the interventions through this lens may indicate that this result was not surprising after all. The interventions used by 6^{th} grade teachers taught students mainly about the importance of believing that intelligence can grow through effort and supporting that with information about neuroplasticity and how the brain continues to grow. However, this and other interventions do not teach students about learning strategies. Thus, it makes sense that students may not be able to act on their newly held beliefs, especially when they face challenges and make mistakes because they are not taught how to. Carol Dweck's recent commentary is as much an indictment on researchers as it is on practitioners (Dweck & Yeager, 2019). There is a saying that goes "what is assessed is valued." Research, and by extension assessment, of growth mindset has been focused on intelligence beliefs, so it is no surprise that practitioners are also focused on the intelligence belief as well. As recent studies and the present study indicate, there may be additional dimensions (e.g., effort, persistence, challenge, mistakes, and learning strategies) that are more nuanced than just a belief (e.g., affect, behavior, cognition), which influence a student's growth mindset and whether they actually act on it.

Implications for Research

First, the results of the structural stage provide supporting evidence that the current items can be improved and that additional research is needed. In terms of the factor analysis results, Factor 1 was composed of items across dimensions that ask about negative emotional responses: "I get frustrated… I give up… I am afraid." One explanation for these items loading onto a common factor was that these items were all negatively oriented. Not only has research shown that negatively oriented items are more difficult to understand (Sherman, 1973; Wason, 1959), it also has shown that they can be related to reading level (Marsh, 1986) which is particularly concerning for this particular 6th grade population. Making all items positively oriented may help improve factor structure and reliability of the MGM measure.

Second, further researcher could examine whether the proposed dimensions are indeed separate and how they are inter-related. For example, the results of the factor analysis showed that Factor 6 was composed of items from separate dimensions. Specifically, the content of the items suggested they may share a common construct of continual or sustained effort. Interestingly, while this construct seems related to persistence, only one item was a persistence item. The other two items were an effort item and learning strategy item. This brings into question whether effort and persistence are two separate dimensions or just two aspects of one effort dimension. Consider the following persistence item that was proposed for the MGM scale: "*When things are hard to learn, it is important to keep trying until you learn it*." It is clear that effort is part of persistence. Similarly, are mistakes and challenge separate dimensions? When something is challenging it is implied that mistakes are more common. One might say mistakes are necessary for challenge. It is important for future research to establish where the line is drawn by first going back to Benson's (1998) substantive phase. If the items are not measuring what is intended, then conclusions and inferences based on them will be faulty.

Third, to guide the creation of the survey, I followed the steps presented by Gehlbach and Brinkworth (2011). The second step was to conduct interviews and focus groups with the population of interest. One of the findings from these interviews was that most students were not familiar with the word “learning strategies.” Students were familiar with common strategies such as studying, flash cards, doing practice problems, etc., but were just not familiar with the overall phrase of learning strategies. This vocabulary barrier posed a problem when writing items for learning strategies and subsequently in the factor analysis results as well. One of the learning strategy items, *I’m not sure how best to learn in school*, did not load onto any factor. Further, the other two learning strategy items loaded onto separate factors. None of the learning strategy items stuck together indicating that the items were not clearly measuring learning strategy and the substantive construct of interest.

Finally, the present research was conducted with only 6th grade students at one middle school, limiting the generalizability of the results. In the future researchers should consider examining students in various grades, with different populations, and in more contexts. It should be noted, however, that the limited population was driven by the by the problem itself. A team of 6th grade math teachers at this particular middle school noticed a problem with their students.

Implications for Practice

This research also has implications for practice. With additional dimensions of growth mindset identified beyond intelligence belief, this suggests that only teaching students about the intelligence belief is not enough for meaningful change to occur in the behavior of students. This was the exact situation that led to this research study. Another example can be seen in a case study from Eskolta (a non-profit educational group that provides K-12 professional development; Podesta, 2015) that captured the steps they found most helpful for teachers if they wanted to create a growth mindset climate in their classrooms that truly impacted students. For example, a teacher they partnered with shared her journey in developing a growth mindset oriented classroom, which took place across four steps. During the first step, the teacher created a lesson to introduce students to intelligence beliefs. In the second step, she followed up using growth mindset language with her students and encouraging them to do the same. In the third step, she recognized that she needed to change her teaching strategies to better support growth mindset concepts (e.g., leading students through discussions on the process of learning, and focusing on the skills that they are developing from a particular assignment rather than the graded outcome). In the final step, the teacher noticed that some of her students still struggled due to one of two reasons. Students either struggled because they lacked learning strategies or because they didn’t see value in what they were learning. Here she recognized that to help both groups of students, she had to instruct them beyond growth mindset beliefs. She either had to help them learn about better learning strategies or to help them find value in the course material.

This example highlights two important takeaways for bringing growth mindset into practice. First, there are many ways growth mindset can be encouraged in students besides an intervention delivered by researchers, as is often the case in the research literature. Second, through the process of this teacher helping her students adopt a growth mindset, she discovered that she couldn’t stop at teaching students about intelligence beliefs. In addition to an initial lesson, the teacher changed her language and teaching strategies, and to help students who still struggled she had to differentiate and provide guidance on using learning strategies or finding value in the course material.

The insight on needing to teach learning strategies has implications for our practitioners. Early in the process (Step 2 of Gehlbach and Brinkworth, 2011), I discovered that students did not understand the phrase learning strategy, even though they could recognize common study strategies. Because of this information, the items written for learning strategy had to avoid that phrasing and ended up not performing as well. While the validity information from the present MGM measure is still emerging, we have evidence from both research and practice demonstrating the importance of learning strategies in students acting on their growth mindset beliefs. Based on that information, I highly recommend to our partner teachers and to education practitioners everywhere that they spend some time talking to students about learning strategies and the process of learning. This may be one of the most important pieces in determining whether a student can actually demonstrate those important underlying behaviors.

Beyond just learning strategies, this study suggests that beliefs on other dimensions also may contribute to a students' overall growth mindset (e.g., how students think about mistakes and challenge). If a student believes that mistakes are an indictment on their intelligence it would be difficult for that student to take risks in school that would lead to greater learning. Additionally, if a student doesn't think challenge is important, then that student will be content to continue with work that may be easy but won't lead to deeper understanding. A great example of this comes from MindsetWorks, a company founded by colleagues of Carol Dweck. One of the free materials available is a growth mindset rubric called the "Effective Effort Rubric." The rubric serves as a selfassessment for students and includes several areas such as: taking on challenges, learning from mistakes, perseverance, and practice and applying new learning strategies, among others.

Next Steps

In summary, there are many directions for future research. First, I would suggest returning to the substantive stage to gain a greater understanding of what and how many dimensions there are to growth mindset. Based on past theorizing and reading of the literature, it seems to be clear that there is more to growth mindset than intelligence belief. However, how the various supporting beliefs of effort, persistence, mistakes, challenge, learning strategies, and perhaps others intertwine and diverge is still not clear. More research on each dimension and supporting evidence of relation to other constructs can help elucidate this.

Second, the MGM measure can be improved. The evidence collected in the present study shows that many of the items were not operating as intended. The immediate areas of concern are with the items loading onto Factors 1 and 6. Additionally, focus should be directed to the dimensions of persistence and learning strategies where the items did not load onto a common factor.

Lastly, as with any line of research, it is always important to gather evidence from other samples. The present study was conducted solely with 6th grade students. While this was largely driven by the problem of practice, it does limit the generalizability of the findings. More research should be conducted with other students in other contexts. In particular, I would argue for more research to be conducted with younger, elementary students.

Returning to the reason why growth mindset research began at THMS middle school, it was because teachers (including 6th grade teachers) indicated that their students had given up on themselves as learners. In the case of the 6th grade teachers, it also indicates that students are leaving elementary school with a fixed mindset about their education and intelligence. A quote from Fredrick Douglass is particularly enlightening, "It is easier to build strong children than to repair broken adults." The earlier we can help students adopt a stronger mindset toward their intelligence and education, the easier the burden will be on future teachers and the more those students will be able to get out of their education experiences.

Appendix A

Student Interview Protocol

Purpose - answer the question "Why are some students growth mindset and other students aren't?" What led to be growth mindset? And what prevents them from shifting to growth mindset if they are still fixed?

Materials and Support for Running Interview

- Room to conduct interviews (talk to teachers 2/1 about this)
- Identify a good timeframe to come interview with students (talk w/ teachers 2/1)
- Ask for performance information on the student
- Have a copy of the script for introducing and closing the interview
- Copy of the interview questions and pen/pencil for taking notes
- If taking handwritten notes, type them up ASAP

Script for Running the Interview

- Email teacher of student you are interviewing the day before your desired time
- Arrive at THMS 15 minutes before the scheduled interview time
- Sign in at the front desk
- Prepare the interview room
    - Make sure it is unlocked
    - Ensure there is enough seating
- Introduce yourself and chit chat with student on the way to the room

- When you get settled in the interview room say the following to set up the interview (it will be more natural if you say something similar without reading directly):

Read statements in bold aloud:

We are meeting with students in 6th grade to get your opinions about 6th grade. There are no right or wrong answers, we just want you to respond openly and honestly because we value what you have to say. Your teachers will not see your responses so this won't affect your grades in any way.

Do you feel comfortable that you can be open and honest? Do you have any questions for me before we start?

I am going to take notes. I want to make sure to capture all of your ideas that you are going to share.

If they say no, is there a way to make them feel more comfortable?

- **The goal is to help improve 6th grade**
- **Everyone's opinion matters, especially students**

During the interview

- Interview should be approximately 15-20 minutes
- Semi-structured protocol.

    - Based on student responses you may find that you need to probe deeper for a student's response.
    - We have a predetermined list of questions, but you might want to re-order.
    - Take notes on how you deviate from the protocol.
- Here is a list of common words or phrases that students have asked questions about along with alternative ways to describe the words:
    - Challenge - difficulty in a task that is stimulating to the one engaged in it
    - Mistakes
    - Learning strategies
    - Others?

Start a timer OR Record start time

First I am going to ask you some general questions about your 6th grade experience so far.

What is your favorite subject in school? Why?

What is your least favorite subject? Why?

Who is your favorite 6th grade teacher? Why?

Now I am going to ask you some questions that will go a little deeper on your experiences in 6th grade.

First, I want to ask you some questions about how much your classes challenge you.

When I say "challenge," what does that mean to you? What other words do you think it is like?

Is challenge a good or bad thing? Why?

Are you challenged in 6th grade classes?

If "Yes," what do your teachers do to challenge you? If "No," what could your teachers do to challenge you?

Now I am going to ask you some questions about the learning strategies you use for schoolwork (in school or at home).

When I say "Learning Strategy," what does that mean to you?

Can you give me an example of one you use? Can you give me any more examples? Do you use different strategies in different class?

The next couple questions will help you think about the strategies you might use in specific situations.

When you are doing school work in class and you get stuck what do you do?

What do you do when you are doing work at home?

Do your teachers give you suggestions for what to do when you get stuck on work in class or at home? If yes, can you give me an example?

Other Questions to ask:

What does studying mean to you

Now I want to ask you some questions about how you view effort.

Some students believe that they are either born smart or stupid and there is nothing they can do to change that (for example, some students believe they were either born good at math or not). What do you believe and why?

Where did you learn this belief?

What do you think your classmates believe? Do all of your classmates believe this? What percent?

What do you think your teachers believe?

What specifically do your teachers do or say that makes you think that?

Capture possible differences between teachers.

Next, I want to ask you some questions about making mistakes.

When I say “mistake,” what does that mean to you? Are there other words that are like it?

Is making mistakes a good or bad thing? Why? *(How do you feel when you make a mistake? When you make a mistake what does it tell you?)*

What is the difference between failing and making a mistake?

Feeling?

Do you get chances in class to make mistakes and learn from them? If yes, can you give me an example? Do you want chances to make mistakes and learn from them?

Finally, I have a couple fun but important questions for you.

Compared to elementary school, do you like middle school more, the same, or less?

If you could king or queen for the day, what would be the one thing you would change about 6th grade to make it better?

Stop timer OR Record stop time

Post-Interview Reactions

Post interview

1. Was the student be comfortable talking?
2. Is the student growth or fixed mindset based on the interview?
3. Did the protocol reveal teacher behavior or other classroom/school practices to promote a growth mindset climate?
4. Was the student share helpful information? Do they say something that is quotable?
5. Was the student honest?
6. How much of the protocol were you able to complete?
7. Would you suggest any changes/adaptations to the protocol/questions?

Appendix B

Multi-dimensional Growth Mindset Measure

1	2	3	4	5	6
Slightly Disagree	Disagree	Strongly Disagree	Slightly Agree	Agree	Strongly Agree

Note. These response options are applied to all items

Theories of Intelligence Scale (Dweck, 1999)

1. (ib1) You have a certain amount of intelligence and you really CANNOT do much to change it.
2. (ib2) Your intelligence is something about you that you CANNOT change very much.

Effort

3. (effort_b) I often work hard in school.
4. (effort_a) If I have to work hard in school, it makes me feel bad about myself.
5. (effort_c) I know I can learn in school if I put in the effort.

Persistence

6. (persist_c) When things are hard to learn, it is important to keep trying until you learn it.
7. (persist_a) I get frustrated if something takes a long time to learn.
8. (persist_b) I give up when things take a long time to learn.

Mistakes

9. (mist_a) I am afraid of making mistakes in school.
10. (mist_c) Making mistakes is an opportunity for learning.
11. (mist_b) I learn a lot from my mistakes or the mistakes of others.

Challenge

12. (chall_a) I like it when school is challenging.
13. (chall_b) I often challenge myself in school.
14. (chall_c) I like challenging material more than easy material.

Learning Strategies

15. (learn_c) I'm not sure how to best learn in school.
16. (learn_a) I get frustrated because I don't know what the best ways are to learn in school.
17. (learn_b) When I get stuck, I try to figure it out before I ask for help.

Manipulation Check

18. How honest were you with each of your answers today?
    a. Never
    b. Part of the time
    c. About half of the time
    d. Most of the time
    e. Always

Appendix C

Results of the four methods for determining the number of factors to retain.

Method 1: K1 Criterion.

Table C1

Eigenvalue results from an initial factor analysis with no factor number specified

	Initial Eigenvalue		
Factor	Total	% Variance	Cumulative %
I	5.67	35.34	35.34
II	1.66	10.36	45.71
III	1.35	8.42	54.13
IV	1.12	7.02	61.15
V	.946	5.91	67.06
VI	.875	5.47	72.52
VII	.712	4.45	76.98
…			

Method 2: Scree Plot.

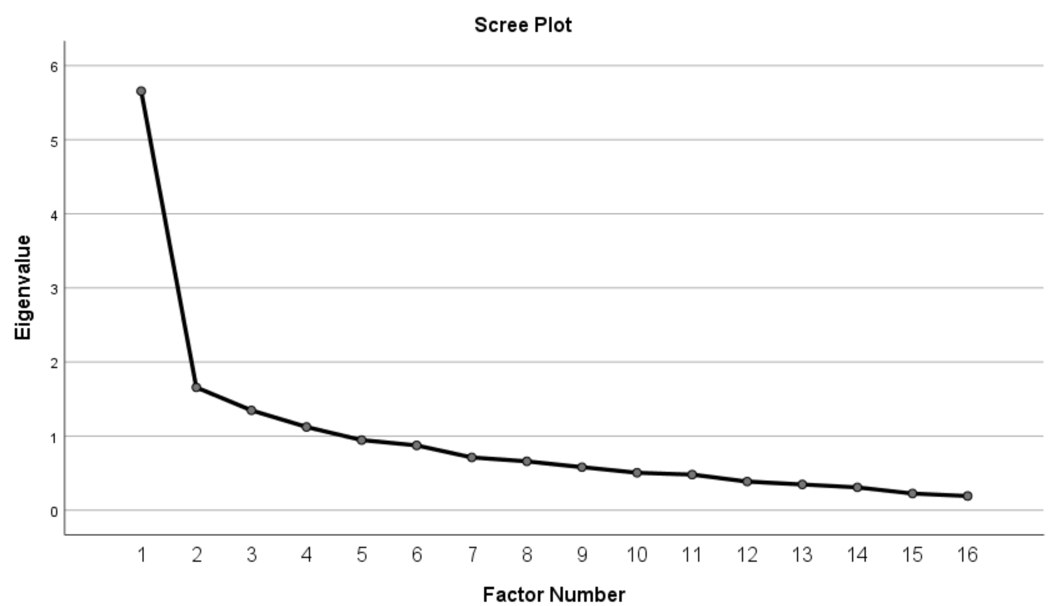

Figure C1. Scree plot obtained from an initial factor analysis with no factor number specified.

Method 3: Parallel Analysis.

Table C2

Random data eigenvalues generated via parallel analysis procedure.

Factor	Eigenvalue	Mean	95th Percentile
1	5.611908	.764129	.888452
2	1.231106	.613584	.744083
3	.884815	.506334	.578122
4	.693676	.419207	.501879
5	.365185	.338016	.412776
6	.302611	.262550	.329647
7	.110711	.184367	.263971
8	.059722	.120616	.189026
9	.016332	.059618	.124175
…			

Method 4: Minimum Average Partial.

Table C3

Average partial correlations to the 4th power obtained from the MAP procedure.

Factor	4th Power
0	.9998
1	.0493
2	.0286
3	.0187
4	.0117
5	.0144
6	.0116
7	.0162
8	.0275
9	.0370
…	

Appendix D

Item level correlations to final grades in all four core classes (English, Math, Science, History).

Table D1

Correlations between MGM items and final grades.

Dimension	English Final	Math Final	Science Final	History Final
ib_1	0.4	0.35	0.38	0.29
ib_2	0.35	0.44	0.43	0.28
effort_a	0.14	0.18	0.21	0.05
effort_b	0.37	0.29	0.26	0.19
effort_c	0.31	0.27	0.29	0.23
persistence_a	0.1	0.23	0.21	0.09
persistence_b	0.27	0.34	0.39	0.29
persistence_c	0.28	0.14	0.28	0.2
mistakes_a	0.06	0.25	0.22	0.13
mistakes_b	0.17	0.08	0.24	0.07
mistakes_c	0.13	0.06	0.19	-0.05
challenge_a	0.38	0.38	0.44	0.16
challenge_b	0.27	0.3	0.29	0.03
challenge_c	0.33	0.34	0.37	0.19
learn_a	0.28	0.31	0.38	0.08
learn_b	0.22	0.14	0.22	0.12
learn_c	0.24	0.23	0.24	0.23

Note. Final = Final grade in class representative of performance throughout the school year, across all four quarters; English *N* = 149; Math *N* = 149; Science *N* = 148; History *N* = 148. We do not have full class schedule information on 6th grade students. In the dataset, there are several students that do not have grades posted in all four of the core classes.

Appendix E

Item level correlations to English and Math SOL scores

Table E1

Correlations between MGM survey items and SOL scores.

Item	English SOL	Math SOL
ib_1	0.5	0.47
ib_2	0.49	0.52
effort_a	0.21	0.26
effort_b	0.17	0.35
effort_c	0.26	0.34
persistence_a	0.25	0.33
persistence_b	0.31	0.37
persistence_c	0.07	0.23
mistakes_a	0.19	0.25
mistakes_b	0.09	0.17
mistakes_c	0.17	0.26
challenge_a	0.29	0.49
challenge_b	0.15	0.36
challenge_c	0.31	0.43
learn_a	0.35	0.42
learn_b	0.17	0.2
learn_c	0.29	0.22

Note. N = 146. SOL refers to the standardized tests taken by k-12 students at the end of the school year. In 6th grade the only SOLs that are taken are Reading (English) and Mathematics.

References

Abd-El-Fattah, S. M., & Yates, G. (2006). Implicit theory of intelligence scale: Testing for factorial invariance and mean structure. *Paper presented at the Australian Association for Research in Education Conference, Adelaide, South Australia.*

American Educational Research Association (2014). *Standards for Educational and Psychological Testing*. DC; American Psychological Association.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113-125. http://dx.doi.org/10.1006/jesp.2001.1491

Bandalos, D. L. (2018) *Measurement theory and applications for the social science*. New York: Guilford Press. https://lccn.loc.gov/2017007303

Barron et al. (2017)

Barron, K. E., Hulleman, C. S., Hartka, T. A., & Inouye, R. B. (under review). Using improvement science to design and scale up psychological interventions in schools: The case of the growth mindset app. https://www.researchgate.net/publication/319377456_Using_Improvement_Science_to_Design_and_Scale_Up_Social_Psychological_Interventions_in_Schools_The_Case_of_the_Growth_Mindset_App

Barron, K. E., & Hulleman, C. S. (2013). *Teacher perceptions of student motivational challenges and best strategies to enhance motivation*. Paper presented in symposium, J. Turner (Chair), Bridging the Theory-Practice Divide: Teacher

Approaches to Motivating Students, at the annual meeting of the American Educational Research Association. San Francisco, CA.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17*(1), 10-17. https://doi.org/10.1111/j.1745-3992.1998.tb00616.x

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*, 246-263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Bettinger, E., Ludvigsen, S., Rege, M., Solli, I. F., & Yeager, D. (2016). Increasing perseverance in math: Evidence from a field experiment in Norway. *Journal of Economic Behavior & Organization*, *146*, 1-15. https://doi.org/10.1016/j.jebo.2017.11.032

Bryk, A. S., Gomez, L., Grunow, A., & Le Mahieu, P. (2015). *Learning to improve: How America's schools can get better at getting better.* Cambridge, MA: Harvard Education Publishing.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. http://dx.doi.org/10.1037/0021-9010.78.1.98

Crandall, V. C., Katkovsky, W., & Crandall, V. J. (1965). Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. *Child Development*, *36*(1), 91-109. https://doi.org/10.2307/1126783

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.

Cronbach, L. J. (1961). *Essentials of Psychological Testing*. New York: Harper & Row. https://doi.org/10.1002/sce.3730350432

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-302. http://dx.doi.org/10.1037/h0040957

Cronbach, J. L. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), *Intelligence: Measurement and theory: Proceedings of a symposium in honor of Lloyd G. Humphreys* (pp. 147-167). Champaign, IL; University of Chicago Press.

De Castella, K., & Byrne, D. (2015). My intelligence may be more malleable than yours: The revised implicit theories of intelligence (self-theory) scale is a better predictor of achievement, motivation, and student engagement. *European Journal of Psychology of Education*, *30*(3), 245-267. https://doi.org/10.1007/s10212-015-0244-y

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology*, *36*(5), 451-462. http://dx.doi.org/10.1037/0022-3514.36.5.451

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087-1101. http://dx.doi.org/10.1037/0022-3514.92.6.1087

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgements and reactions: A world from two perspectives. *Psychological Inquiry*, *6*(4), 267-285. https://doi.org/10.1207/s15327965pli0604_1

Dweck, C. S. Tenney, Y. & Dinces, N. (1982). Implicit theories of intelligence as determinants of achievement goal choice. Unpublished manuscript.

Dweck, C. S., & Bempechat, J. (1983). Children's theories of intelligence. In Paris, G. O. & Stevenson, H. (Eds.) *Learning and motivation in the classroom* (pp. 239-256). London, ENG: Routledge. https://doi.org/10.4324/9781315188522

Dweck, C. S., & Henderson, V. L. (1989). Theories of intelligence: Background and measures. *Paper presented at the Biennial Meeting of the Society for Research in Child Development, Kansas City, Missouri*.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation psychology. *Psychological Review, 95*, 256-273. http://dx.doi.org/10.1037/0033-295X.95.2.256

Dweck, C. S., & Reppucci, D. (1973). Learned helplessness and reinforcement responsibility in children. *Journal of Personality and Social Psychology*, *25*(1), 109-116. http://dx.doi.org/10.1037/h0034248

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science*, 1-16. https://doi.org/10.1177/17456916188041

Dweck, C. S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology*, *31*, 674-685. http://dx.doi.org/10.1037/h0077149

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. New York: Psychology Press.

Eisinga, R., Grotenhuis, M., Pelzer, B. (2012). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*(1), 5-12. http://dx.doi.org/10.1037/0022-3514.54.1.5

Faria, L., & Fontaine, A., (1997). Adolescents' personal conceptions of intelligence: The development of a new scale and some exploratory evidence. *European Journal of Psychology of Education*, *12*(1), 51-62. https://doi.org/10.1007/BF03172869

Froehlich, L., Martiny, S. E., Deaux, K., Goetz, T., & Mok, S. Y. (2016). Being smart or getting smarter: Implicit theory of intelligence moderates stereotype threat and stereotype lift effects. *British Journal of Social Psychology*, *55*, 564-587. https://doi.org/10.1111/bjso.12144

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology*, *15*(4), 380-387. https://doi.org/10.1037/a0025704

Gripshover, S., Beaubien, J., Romero, C. L., Walton, G., Dweck, C.S., Yeager, D., & Paunesku, D. (2017). *The growing impact of learning mindset interventions among community college students*. Working paper. Stanford, CA: Project for Education Research That Scales (PERTS).

Henderson, V. L., & Dweck, C. S. (1989). Predicting individual differences in school anxiety in early adolescence. *Paper presented at the Biennial Meeting of the Society for Research in Child Development, Kansas City, Missouri*.

Hong, Y., Chiu, C., Dweck, C. S., Lin, D. M., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*(3), 588-599. http://dx.doi.org/10.1037/0022-3514.77.3.588

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527-535. http://dx.doi.org/10.1037/0033-2909.112.3.527

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

Kosovich, J. J., Hulleman, C. S., & Barron, K. E. (2019). Measuring motivation in educational settings: A Case for pragmatic measurement. In K. A. Renninger & S. E. Hidi (Eds.), *The Cambridge Handbook on Motivation and Learning* (pp. 713-738). New York, NY: Routledge. https://doi.org/10.1017/9781316823279.030

Levy, S., & Dweck, C. S. (1997). *Implicit theory measures: Reliability and validity data for adults and children*. Unpublished manuscript, Columbia University, New York

Marsh, H. (1986). Negative item bias in rating scales for preadolescent children: A cognitive developmental phenomenon. *Developmental Pscyhology*, 22, 37–49. http://dx.doi.org/10.1037/0012-1649.22.1.37

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain: School and corporate applications.* 3rd Ed. New York: Springer.

Miller, R. G. (1981). *Simultaneous statistical inference*. 2nd Ed. New York; Springer

Nunnally, J. C., Bernstein, I. H. (1994). *Psychometric Theory*. 3rd Ed. New York: McGraw-Hill.

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science*, *26*, 784-793. https://doi.org/10.1177/0956797615571017

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* Thousand Oaks: Sage Publications.

Podesta, P. (2015). *An exemplary growth mindset classroom: A case study of one teacher's journey in changing student mindsets.* New York: Eskolta.

Romero, C., Master, A., Paunesku, D., Dweck, C. S., & Gross, J. J. (2014). Academic functioning in Middle School: The role of implicit theories. *Emotion*, 227-234. https://doi.org/10.1037/a0035490

Seligman, M. E., Maier, S. F., & Greer, J. (1968). Alleviation of learned helplessness in the dog. In Keehn, J. D. (Ed.) *Origins of Madness: Psychopathy in Animal Life* (1979; pp. 401-409). Elsevier. https://doi.org/10.1016/B978-0-08-023725-1.50049-5

Sherman, M. A. (1973). Bound to be easier? The negative prefix and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *12*, 73–84. https://doi.org/10.1016/S0022-5371(73)80062-3

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mindsets important to academic achievement? Two meta-analyses. *Psychological Science,* 1-23. https://doi.org/10.1177/0956797617739704

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, *52*, 92–107. https://doi.org/10.1080/17470215908416296

West, Finch, & Curran (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA, US: Sage Publications, Inc.

Wormington, S. V., Kosovich, J. J., Hulleman, C. S., Howard, D., Phelps, J., & Lee, M. (under review). Is effort enough or do I need strategies, too? Assessing the fidelity and efficacy of a growth mindset intervention in developmental mathematics.

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... & Trott, J. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*, 374-391. http://dx.doi.org/10.1037/edu0000098

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

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of factors to retain. *Psychological Bulletin*, *99*(3), 432–442. https://doi.org/10.1037/0033-2909.99.3.432