San Jose State University

SJSU ScholarWorks

Summer 2021

The Role of Self-Efficacy and Academic Mindset on Middle School Math Achievement for At Promise Youth: An Explanatory Study

Michael Lockwood Mansfield
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_dissertations

THE ROLE OF SELF-EFFICACY AND ACADEMIC MINDSET ON MIDDLE
SCHOOL MATH ACHIEVEMENT FOR AT PROMISE YOUTH: AN
EXPLANATORY STUDY

A Dissertation

Presented to

The Faculty of the Educational Doctoral Program in Educational Leadership

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

by

Michael Mansfield

August 2021

The Designated Dissertation Committee Approves the Dissertation Titled

THE ROLE OF SELF-EFFICACY AND ACADEMIC MINDSET ON MIDDLE
SCHOOL MATH ACHIEVEMENT FOR AT PROMISE YOUTH: AN
EXPLANATORY STUDY

by

Michael Mansfield

APPROVED FOR THE EDUCATIONAL DOCTORAL PROGRAM IN
EDUCATIONAL LEADERSHIP

SAN JOSÉ STATE UNIVERSITY

August 2021

Brent Duckor, Ph.D.   Department of Teacher Education

Roxana Marachi, Ph.D.   Department of Teacher Education

Cheryl Roddick, Ph.D.   Department of Mathematics and Statistics

ABSTRACT

THE ROLE OF SELF-EFFICACY AND GROWTH MINDSET ON MIDDLE SCHOOL
MATH ACHIEVEMENT FOR AT PROMISE YOUTH: AN EXPLANATORY STUDY

by Michael Mansfield

California's accountability system has blended its focus on noncognitive factors of achievement as well as evidence of academic growth. School districts are becoming increasingly interested in developing noncognitive factors in their students to help increase student achievement. With the wide variety of mandates and responsibilities schools are shouldering, it will be important to continue developing methods to increase all school accountability metrics in relation with each other. This research aims to explore the relationship between noncognitive factors and mathematics achievement as measured by the Northwest Evaluation Association's Measure of Academic Progress assessment (MAP). This study surveyed 8th graders throughout a school district to measure their ratings of self-efficacy and growth mindset, two major noncognitive factors found in the research to be major drivers of student achievement. During this research study I developed a new instrument, the Self-Efficacy and Academic Mindset (SEAM). The survey instrument was adapted from the Mindsets Essential Skills and Habits survey (MESH), a survey used by the CORE districts of California to measure and evaluate the noncognitive factors of their students. The survey was conducted in conjunction with the fall 2020 administration of the MAP assessment.

DEDICATION

Give thanks to the Lord, for he is good; his love endures forever. Thank you to my parents for supporting my goals and giving me every opportunity to succeed. To my friends and family, thank you for your encouragement and belief that I could do this.

Thank you to Dan Ordaz and Jackie Guevara for taking a shot on a young teacher who didn't know much and for helping me see that I was an educational leader. Thank you to my UPA family for teaching me early in my career how truly special a school community can be. Thank you to all the great colleagues who I have had so much fun working so hard with: Jeremy Powell, Andrew Yau, Dustin Mones, and Ivan Alcaraz you all taught me so much. Ann Kilty, thank you for picking me up when I wanted to quit.

Thank you to every student I have ever had the pleasure to teach, council and support. You continue to teach me every day and you make me appreciate how blessed I am to be an educator, the best job in the world.

Dr. Duckor, thank you for pushing me to the max and then a little further. You helped support me to do research I would never imagined I was capable of doing. Dr. Marachi and Dr. Roddick your guidance and support during this process were also much appreciated. There were also a lot of teachers and coaches along the way who inspired me along the way but one in particular who taught me the greatest lesson an educator can learn. It's not about what you teach your students but how you show them you care. Thank your Mr. Mattingly for not kicking me out of the room when you picked up one of my broken classmates.

BCE, it was a fun ride, we couldn't have had a better group, we were the castaways from Lost without the death, destruction or disappointing ending.

I would not be here without my wife Deauna, who has steadfastly endured a lot of late work nights, who has always been able to pick me up when I didn't think I could keep going, and who has encouraged me to pursue my passion. To my daughters Bethany, Adah, and Abi, thanks for being patient with all my "great work" to be done and being the most wonderful daughters a dad could ask for.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CAASPP – California Assessment of Student Performance and Progress
CAQDAS – Computer Assisted Qualitative Data Analysis Software
CGI – Conditional Growth Index
CGP – Conditional Growth Profile
CORE districts – Fresno Unified, Garden Grove Unified, Long Beach Unified, Los Angeles Unified, Sacramento City Unified, San Francisco Unified, Santa Ana Unified
DIF – Differential Item Functioning
DOK – Depth of Knowledge
EAP – Expected A Posterior
IRT – Item Response Theory
LCAP – Local Control and Accountability Plan
LCFF – Local Control Funding Formula
MAP – Northwest Evaluation Association's Measure of Academic Progress Growth Assessment
MESH – Competencies that demonstrate mindsets, essential skill development and habits
NCLB – No Child Left Behind Act
NCRTI – National Center on Response to Intervention
NWEA – Northwest Evaluation Association
PACE – Policy Analysis for California Education Group
RIT – Rasch Unit
SBAC – Smarter Balanced Assessment Consortium
SEAM – Self-Efficacy & Academic Mindset Survey
SPSS – Statistical Product and Service Solutions
STAR – Standardized Testing and Reporting
UDL – Universal Design for Learning
URCML – Unidimensional Random Coefficient Multinomial Logit
VAM – Value Added Modeling

Chapter I: Introduction

In a nationwide survey of high school dropouts, 69 percent said that school had not motivated or inspired them to work hard (Bridgeland et al., 2006). Lack of motivation not only prevents students from completing high school but also impacts their ability to perform their best. Over a decade removed from the No Child Left Behind Act (NCLB), and with the new accountability frameworks adopted by the state of California, educators are held accountable not only for test scores and academic proficiency, but for a host of other factors as well. In California's new system there has been a redirection toward a more holistic view of accountability, with ten outlined priorities that go beyond just assessment. Many superintendents and school leaders have applauded the change (Fullan & Rincon-Gallardo, 2017). Among these ten priorities, student achievement, student engagement, school climate, parental involvement, provision of basic services, curriculum access, and implementation of state standards are all monitored and evaluated. Yet, the major measure for academic achievement is still end-of-year testing. And, while the new California Assessment of Student Performance and Progress (CAASPP) is more rigorous than the Standardized Testing and Reporting (STAR) assessments they replaced, they still do not provide students or teachers with meaningful opportunities to observe growth within the school year or to inform their classroom instruction. A report released recently from the Policy Analysis for California Education Group (PACE) alludes to another concern:

> The state has chosen a simplistic "change" measure by merely taking the
> difference between this year's scores and last year's scores on each outcome.
> This approach suffers from many problems, not the least of which is that it
> does not adjust for the fact that these are different students being compared to

> one another (i.e., there are "cohort effects"). Especially for test scores, where there is a wealth of knowledge about the best ways to construct accountability system growth measures, there is no reason for the state to choose the approach it did. The state should choose a more appropriate growth measure, such as a two-step value-added model. (Polikoff, 2019)

The state of California is trying to recognize a more holistic view of student needs while still holding schools accountable, but the core academic measure is still a work in progress.

If educators are to continue using standardized tests as one of the state's major metrics for academic accountability, then we must find ways to utilize standardized test results to inform and improve classroom instruction from both the teacher and student perspectives. This should also include making connections between test performance and the noncognitive factors that affect achievement. These noncognitive factors lie behind some of those nine "non-academic" factors tracked by the state. Much research has been done into various noncognitive factors that inform student achievement but there are gaps in the understanding of how these noncognitive factors interact with each other, and how they directly impact student achievement (Farrington et al., 2012). Today there exists a striking opportunity, with the state's increased focus on tracking and developing noncognitive indicators for student success, to explore how these indicators may directly lead to student achievement, and vice versa.

Statement of the Problem

Math achievement is a major determinant of college and career readiness both in the state of California and nationally. Unfortunately, current math achievement levels suggest that the system is not preparing students for college and career readiness. Statewide in

California, only 39.73 percent of students met or exceeded standard in mathematics, across all grade levels and subgroups tested (California Department of Education, 2020). Over half of the students taking this state assessment have not met or exceeded the standard; in the district where this study was conducted one can see the issue even further magnified. Table 1 contains a comparison of state test scores in the target district and across the state.

Table 1

CAASPP Mathematics Achievement 2018-2019

Grade	Target District Meet or Exceed Standard	California Meet or Exceed Standard
Grade 3	30.93%	50.22%
Grade 4	29.41%	44.94%
Grade 5	22.32%	37.99%
Grade 6	18.87%	38.52%
Grade 7	20.37%	37.84%
Grade 8	20.41%	36.63%
Grade 11	20.54%	32.24%
All Grades	23.54%	39.73%

As Table 1 illustrates, the systematic and consistent score reports of underperformance in meeting or exceeding mathematics standards make it more likely that these students will not be able to be college and career ready. Closer examination of two of the largest subgroups in the district where this study was conducted highlight an even more troubling pattern; Table 2 contains test score results for Latinx students, and Table 3 contains test scores results for ELL students.

Table 2

CAASPP Mathematics Achievement Among Latinx Students

Grade	Target District Meet or Exceed Standard	California Meet or Exceed Standard
Grade 3	25.59%	40.02%
Grade 4	24.55%	33.68%
Grade 5	17.67%	26.39%
Grade 6	12.57%	26.81%
Grade 7	12.47%	25.31%
Grade 8	13.22%	24.15%
Grade 11	14.58%	20.27%
All Grades	17.38%	28.05%

As Table 2 reports, mathematics achievement levels at all grade levels lagged behind that of the state within the Latinx subgroup in 2018-19.

Table 3

CAASPP Mathematics Achievement Among English Learners 2018-2019

Grade	Target District Meet or Exceed Standard	California Meet or Exceed Standard
Grade 3	14.08%	24.58%
Grade 4	24.55%	17.76%
Grade 5	6.82%	9.59%
Grade 6	.75%	7.5%
Grade 7	1.06%	7.0%
Grade 8	.60%	5.96%
Grade 11	1.15%	5.01%
All Grades	7.49%	12.58%

As Table 3 reports, mathematics achievement levels at all grade levels, with the exception of the 4th grade lagged behind that of the state within the ELL subgroup in 2018-19.

This level of math under-preparedness is further corroborated by other testing data used by the district where this study was conducted. According to the Northwest

Evaluation Association's (NWEA) Measure of Academic Progress Growth Assessment (MAP), 60 percent of the incoming 9th grade students are more than two years below grade level in mathematics (Target District, 2019). Two major challenges are highlighted by these sets of data. First, many students are arriving at high school in this district with a mathematics readiness level that exceeds the capacity of the system to support. Second, the readiness deficit in many cases is so large, that the high schools have struggled to raise their student's achievement levels to meeting or exceeding standard after three years of working with them. Recent target district data also demonstrates that some limited gains have been made at the elementary level in both reading and math across the district, but those gains have not carried over to the middle school level. During the 2019-2020 school year, almost every elementary school and grade level tested within the district experienced low math achievement levels according to MAP (as measured by Rasch Unit, RIT score) but experienced far above average growth (as measured by conditional growth percentile). The overall average growth rate of all elementary students in the target district was in the 68th percentile nationally for mathematics. The middle schools in the target district tell a different story; achievement levels are similarly low to the elementary schools, however, far less growth is occurring at the middle school level. The overall average growth rate of the middle schools in math was in 46th percentile (Target District, 2019). Elementary students are below grade level but the achievement gap is narrowing; unfortunately, these gains have not carried over into the middle schools.

This district and its schools are emblematic of a widespread pattern found throughout the state of California. Many schools and districts have large numbers of students

operating at a low achievement level, creating the potential for high growth, but these schools and districts are challenged to realize this opportunity due to poverty, systematic racism, and institutional barriers. The growing interest in noncognitive indicators likely comes from the assumption that these indicators are more likely to be changed through intervention than the various social, economic, and political forces that impact the students.

This disparity in the test scores raises the question: what is going on at the middle school level in the target district? The rising academic gains secured at the elementary level that sharply diminish at the middle school level suggest that the middle school is an environment ripe for investigation. In the target district, the reality is that the sudden lack of academic growth at the middle school level cannot simply be explained away by socio-economic factors or the home environment. The same students who did not perform at the middle school level *did* perform at the elementary school level. Several districts within the state of California, collectively known as the CORE districts, have begun to focus on noncognitive factors as an area of study to help understand student levels of achievement and help monitor continuous school improvement. The CORE districts are currently using the MESH survey as one of their primary instruments for assessing four noncognitive indicators: self-management, self-efficacy, growth mindset, and social awareness. Their justification for using these noncognitive factors is their belief that they show the most promise for meaningfulness, measurability, and malleability (Transforming Education, 2016). Unfortunately, while valid and reliable instruments

exist to detect these noncognitive indicators within students, they are not often systematically connected to standardized testing or correlated to those scores.

Significance of the Problem

Various research has shown the potential significance of noncognitive indicators in shaping human outcomes in health, social behavior, and labor markets with factors that could not be measured by typical cognitive test scores (Bowles, 1976; Farkas, 2003; Heckman et al., 2006). This study will explore the effect of providing students with an opportunity to study their own academic growth data, specifically their personal growth on the MAP Growth, and monitor how this impacts their academic mindset, motivation, and self-efficacy. In the average classroom, much of a student's perception about their performance in that class is shaped by the feedback and grades provided by the individual teacher. That same classroom is also impacted by the broader school climate, which is influenced not only by teachers, but by administrators, counselors, and support staff as well. Standardized testing results can be impacted by the same factors; however, the MAP Growth exam will give the students an opportunity to see their academic growth relative to their peers both locally and nationally. The potential of seeing the relative growth could demonstrate to students that those gains are substantial. If it can be shown that these noncognitive indicators can both help predict future academic achievement and be increased themselves by that same academic achievement, it may be possible to create a positive feedback loop and make standardized testing something more useful to both the students and teachers in the classroom.

There is tension within our educational system regarding how assessment data are used—and these concerns from parents, students and educators are not unfounded. Misuse of assessment data and the accountability regime of the NCLB era damaged the educational system. Standardized testing data have long been used to make arguments about the lack of preparedness of students, declining performance of teachers, and have been used to justify reforms that weren't always designed to actually improve public education (Berliner, 2013; Berliner, 2014; Berliner & Biddle, 1996b; Glass, 2008; Nichols & Berliner, 2007). However, if educators could continue to evolve the use of assessment data—shifting away from a focus on summative end-of-year testing and toward formative interim assessments that highlight and demonstrate growth—we may be able to make standardized assessments more useful for providing feedback on student progress as well as for improving student academic mindsets, motivation, and self-efficacy. A large body of research has demonstrated the significance of the effect a student's academic mindset has on their academic performance. Students who believe their increased efforts result in increased competence and ability have been shown to have better academic performance (Cury et al., 2006; Dweck & Leggett, 1988). Students' self-efficacy has also been shown to be a major contributing factor to their levels of perseverance, effort, and persistence (Bandura, 1986; Bandura & Schunk, 1981; Bouffard-Bouchard, 1990; Lent et al., 1984; Pajares, 1996; Schunk & Hanson, 1985). The challenge this study will attempt to address is whether or not standardized assessments can be framed and used in such a way that students see them as an

opportunity to demonstrate and evaluate their progress, fueling their motivation and efforts at academic growth.

Standardized testing is not likely to go away. As such, we must find ways to frame its use responsibly, helping students see their growth and progress. Research suggests that growth mindset can boost student persistence, develop healthy levels of self-efficacy, and motivate students (Claro et al., 2016; Dweck, 2008; Ehrlinger et al., 2016; O'Rourke et al., 2014). According to Dweck (2008), "in a growth mindset, people believe that their most basic abilities can be developed through dedication and hard work—brains and talent are just the starting point. This view creates a love of learning and a resilience that is essential for great accomplishment" (p. 210). Perhaps test scores could be used to help boost this mindset.

Purpose of the Study

The purpose of this study is to determine, at the middle school level, the impact of a student's academic mindset and self-efficacy on their math achievement. Furthermore, this study will evaluate the impact of presenting students with their own academic growth data from norm-referenced interim standardized assessments on academic mindset and self-efficacy. This type of feedback could be particularly valuable for the student whose personal or schooling context mask the reality that their growth is more pronounced than they perceive.

Research Questions

The research questions that were addressed in this study are as follows:

- RQ1: What is the relationship, if any, between noncognitive factors and mathematics achievement for at-promise middle school students?
  - RQ1a: What is the relationship, if any, between self-efficacy (SE) and mathematics achievement for at promise middle school students?
  - RQ1b: What is the relationship, if any, between academic mindset (AM) and mathematics achievement for at promise middle school students?
- RQ2: What factors predict math achievement?
- RQ3: How reliably does the SEAM indicator measure levels of self-efficacy and academic mindset? What is the evidence for the validity of SEAM scores to measure students?

Definition of Terms

The following terms will be used throughout this study.

Academic Mindsets

Academic mindsets come from the beliefs, attitudes and perception in one's self in relation to academic achievement and performance. Literature in this area sometimes divides this construct into 4 main domains: 1) I belong in this community, 2) my ability and competence grow with my effort, 3) I can succeed at this, and 4) this work has value for me (Farrington et al., 2012). For the purposes of this dissertation, academic mindset will refer to the domain that ability and competence grow with effort, and the domain that one can succeed at an academic task is referred to as growth mindset (Dweck, 2008).

At Promise Students

The Every Student Succeeds Act of 2015 gave local schools more control of how they spent money and intervened on behalf of "at risk" students, which they define as those who are economically disadvantaged students, students from minority ethnic groups, children with disabilities, and English language learners (Darling-Hammond et al., 2016). In 2019, California passed AB 419, which removed the title "at risk" and replaced it with "at promise."

Attribution Theory

One explanation use to understand behavior comes from attribution theory, "Attribution theory deals with how the social perceiver uses information to arrive at causal explanations for events. It examines what information is gathered and how it is combined to form a causal judgment" (Fiske & Taylor, 1991).

Conditional Growth Index & Condition Growth Percentile

The Northwest Evaluation Association's (NWEA) technical report states that to accurately measure growth, a student must be observed on two or more occasions and each observation must accurately measure performance on a common underlying developmental construct. Two main growth measures are generated from the MAP Growth assessment. The Conditional Growth index (CGI) represents the relative growth of a student compared to their peers; a score of zero means they grew comparatively to their peers, a positive score indicates atypically high academic growth, and a negative score indicates atypically slow academic growth (Northwest Evaluation Association [NWEA], 2019). The Conditional Growth Percentile (CGP) represents the percentile ranking of the CGI. These peer group norms were studied by NWEA and the high

degrees of marginal reliability across the grade levels suggest the assessment is testing

what it was designed to as seen in Table 4.

Table 4

MAP Assessment's Marginal Reliability of RIT Scores (Mathematics)

Grade Level	6	7	8	9	10	11	12
Reliability	0.973	0.976	0.977	0.976	0.978	0.981	0.982
N	68,842	63,735	60,095	36,949	29,601	15,745	7,695

(NWEA, 2019)

When examining by instructional area we can further see the high degrees of reliability of

the MAP assessment, as seen in Table 5.

Table 5

MAP Assessment's Reliability By Instructional Area

State	Algebraic Thinking	Number & Operations	Measurement & Data	Geometry	Real & Complex Number Systems	Statistics & Probability
CA	0.976	0.977	0.976	0.978	0.981	0.982

Note. N=547,912. Adapted from "MAP Growth Technical Report," by NWEA, 2019.

Furthermore, the MAP assessment also has a strong degree of concurrent validity with

the California Smarter Balanced Assessment Consortium (SBAC) assessment, as seen in

Table 6. This strong concurrent validity not only suggests the tests are aligned but opens

up the possibility for the argument that the MAP Growth assessment could be used

instead of the SBAC assessment; however, it needs to be conceded that only the SBAC

tests writing skills directly.

Table 6

Concurrent Validity of MAP and SBAC Mathematics Tests in California

Test		Grade 6	Grade 7	Grade 8
Spring 2015	r	0.89	0.87	0.85
	N	7,042	6,141	5,625

Note. Adapted from "MAP Growth Technical Report," by NWEA, 2019.

Noncognitive factors

An incredibly broad term, noncognitive factors is the term developed to distinguish between the core academic skills and processes one needs to learn to achieve in school (cognitive factors) and the strategies, attitudes, and behaviors one may possess that may drive academic achievement as well (Farrington et al., 2012). Various researchers have pointed out the somewhat false dichotomy between these two constructs; as Borghans et al., 2008 noted, "few aspects of human behavior are devoid of cognition" (p. 974). In this study, particular focus was placed on two noncognitive factors: self-efficacy and academic mindset.

Northwest Evaluation Association's, Measure of Academic Progress Growth Exam (MAP Growth)

The MAP Growth assessment is an interim computer adaptive test, designed to be given 2-3 times a year and track a student's growth and achievement levels as norm referenced to millions of other students across the United States. The stated design purpose of the exam is to create an assessment that dynamically adjusts to individual student achievement levels, provide performance and growth summaries, connect across grade levels (providing an ability to track and understand a student's performance as they

progress through grade levels), and help teachers plan instruction based on targeted growth goals (NWEA, 2019).

Rasch Unit (RIT) Scales

The MAP Growth assessment generates a RIT score. The RIT scale is derived from Item Response Theory (IRT) which describes the relationship between item characteristics and student achievement (Lord, 1980; Lord & Novick, 1968; Rasch, 1980). The RIT score generated in the MAP Growth exam is based on a one-parameter Rasch IRT model that estimates the probability that a student with a particular achievement score will correctly answer a test item of particular difficulty (NWEA, 2019). For the purposes of this dissertation, any reference to RIT scores is a reference to the academic achievement levels as evidenced by the MAP assessment.

Self-Efficacy

Self-efficacy is an individual's belief of whether or not they have the capability to succeed at a particular task in the future (Bandura, 1997, 2001). In the literature, self-efficacy and self-concept are often paired under the larger grouping of self-perceptions. Self-concept evaluates an individual's general feelings about their capability based on past performance, but self-efficacy is a measure of future expectations pertaining to one's capability for future performance.

Site Selection and Sample

The target location for this study was a school district in a rural community in Northern California. The district's middle school population is approximately 85% Latinx, 12% white, 85% socio-economically disadvantaged, 13% homeless, 38% English

learner and 16% students with disabilities. The target classes for the study were 8th grade math classes.

Scope and Limitations of the Study

The study took place over the course of one academic semester. I followed a group of 8th grade math students through MAP Growth assessment administrations in the fall of 2020. The study used a quantitative approach employing the SEAM survey, given in conjunction with the student's administration of the MAP Growth assessment. The purpose of administering the survey in conjunction with the MAP Growth assessment was to explore the relationship between students perceived levels of academic mindset or self-efficacy may relate to their academic performance as measured by the MAP Growth assessment.

It has been argued that self-efficacy measurements are most reliable when the task used to test the self-efficacy is closely aligned in timing with the future related task and the measurement have a high degree of specificity (Bandura, 1997). However, a natural consequence of this specificity is that the results become less generalizable the more specific the testing task becomes. It has also been argued that self-efficacy is difficult to measure; critics have challenged the validity and reliability of imprecise definitions within the instruments' testing self-efficacy (Caprara et al., 2003; Choi et al., 2001; Tschannen-Moran et al., 1998; Tschannen-Moran & Hoy, 2001). Subject students have had uneven backgrounds in elementary and middle school regarding knowledge of what the test means; teachers in different classes or at different schools may have put more or

less emphasis on explaining the test and scores to students before I worked with them This could certainly impact my results.

Assumptions, Background, and Role of the Researcher in the Study

Informing this study is my belief that a primary motivator of human accomplishment is self-efficacy. Bandura (1986) argued, "People who regard themselves as highly efficacious act, think, and feel differently from those who perceive themselves as inefficacious. They produce their own future, rather than simply foretell it" (p. 395). I believe academic mindset is key to driving accomplishment in any field, including education. In my first teaching job at a brand-new charter school, I saw first-hand the importance of self and collective efficacy as they were what I perceived as the fuel that drove the school's growth and success.

I do believe that all students are capable of experiencing the phenomenon of self-efficacy, while I also acknowledge that a host of factors in the control of the student and outside of the control of the student will certainly impact a student's ability to experience self-efficacy. Socio-economic status, family upbringing, cultural values, the environment, and health are just a few of the many factors that could certainly impact a student's self-efficacy and academic mindset.

My research into this topic also comes from a practical and functional mindset. Standardized testing is likely not going to be removed completely from our educational system due to strong political interest in maintaining some version of standardized testing. I am approaching this research with the outlook that if it cannot be removed from our system, it ought to be made more impactful for students and teachers to improve

16

student learning outcomes. My research is aimed at attempting to posit that specific types of standardized testing results framed in a growth minded context could have beneficial self-efficacy strengthening properties that could encourage future further growth. Based on personal experience and research, I am not convinced this would be the only or most effective way to increase this self-efficacy, but I am interested in exploring whether standardized testing, which is fairly loathed across many corners of the educational system, could potentially have value if used more intentionally.

As an assistant principal in a high school whose district is placing focused emphasis on the MAP Growth assessment and the scores being generated, I know that my positionality and the positionality of my district will have an effect on the research. Some of the students involved in the research may have had more extensive discussions than others about the test scores, and therefore may be more or less interested in discussing them in greater depth with me. For teachers who have less faith or interest in the information being captured in the MAP growth assessment, my work may be seen as a distraction from the more legitimate work they wish to focus on, or even as a waste of their time.

Chapter II: Literature Review

This research study was developed through an examination of related literature. In this literature review, I first examined social cognitive theory (Bandura, 1986) and human agency and how those frameworks apply to learning and assessment. Next, I explored academic achievement within the context of California and the Local Control and Accountability Plan (LCAP) and Local Control Funding Formula (LCFF); also, noncognitive factors and their measurement and use within accountability measures in California were examined. The purpose of reviewing California's new academic accountability systems was to highlight potential gaps in practice, namely that our academic assessments still only measure end-of-year growth and do not provide the instructional assessment data teachers and students need to monitor and develop growth throughout the year. I then further examined two noncogntive factors, growth mindset and self-efficacy. These two factors have been particularly focused on throughout the broader research and more specifically in California's work to use noncognitive factors in school improvement and accountability measures. Next, I explored in the literature reviewed the history of high stakes testing with the purpose of documenting the myriad reasons why standardized testing has sometimes been faced with animosity by some in the education profession and how this may have led to a gap in the practice and research. Lastly, I explore the MAP Assessment to establish its validity, reliability, and potential opportunity as a standardized assessment that could be used to bridge noncognitive factors and achievement.

Social Cognitive Theory & Human Agency

When asked "What is the role of school and teachers?" Caleb Carman, an 11th grader

from New York, responded, "The role of education and the role of teachers is to

empower students not just to do what they want, but to make mistakes. The more often

you make mistakes, the more likely you will be to do something important" (Slapik,

2017). In this student's understanding, the role of school and teachers is empowerment.

To have agency is to believe that one has the power to shape their life, environment, or

events by their own actions or influence; in this view, people make their life, their life

does not make them (Bandura, 2006, 2008). The theoretical framework of social

cognitive theory proposes how this empowerment and agency is taught; human

development can be directly attributed to an individual's observations and interactions

with their environment, experiences, and social interactions. In the social cognitive theory

view, human development is often influenced by social factors and social factors can

influence how one perceives their environment, experiences, and social interactions in

reciprocal fashion. Cognition often directly influences human motivation, affect, and

action; these same forces also directly influence cognition (Bandura, 2012). Social

cognitive theory was a rejection of the behaviorist notion that human development was

the result of responses to the environment without cognition or interpretation to those

environmental stimuli through the lens of social experience.

Throughout the development of social cognitive theory debate existed about whether

the antecedent to human development was driven by the environment, experiences, and

social interactions of the individual or whether the interpretations of the environment,

experiences, and social interactions of the individual were the antecedent of human development (Bandura, 2012). The question debated here is an important one, particularly to educators. To put the debate more simply, this debate wrestles with the fundamental question of whether the environment drives human development or if human development is driven by the interpretation of that environment. For education to have a purpose, we must believe that it is interpretation of the environment, experiences, and social interactions that is driving human development because interpretation is not predetermined—it can be taught and learned. However, we cannot ignore the simultaneous reality that an individual's interpretation of their environment, experiences, and social interactions may be driven by factors they believe to be deterministic. To put it another way, it would likely not be very difficult to find a person who would argue that race, class, and/or gender do not determine one's educational outlook; it also would not be difficult to find someone who would argue the exact opposite. Bandura's argument was that these simultaneously existing contradictory outlooks could exist because human functioning is derived from the dynamic interaction of social interactions and behavioral and environmental determinants that do not play out the same for each individual; he called this interplay triadic reciprocal causation (Bandura, 1978). Implicit in Bandura's explanation is the acknowledgement that our environment certainly impacts our development, but that our interpretation of that environment is just as important.

Implications of Social Cognitive Theory and Human Agency on Learning

There are a few broadly acknowledged intersections between human agency and learning that are mediated through a social cognitive theory lens. Donovan et al. (1999)

had three key findings and four practical implications for classroom practice; two of these key findings and implications sit at the intersection of how to develop human agency and learning in the classroom context. The first key finding is that students come to the classroom with preconceptions and misunderstandings; the notion that students arrive in classrooms as empty vessels ready to be filled by their teachers is not supported by the research (Donovan et al., 1999). Students' social interactions, experiences and environment begin to help them develop some sense of the world at an early age. Preschoolers have been shown to develop understandings of the world and phenomenon around them (Wellman, 1992). Students come into classrooms with ideas and conceptions that relate not only to knowledge and academic skills but to noncognitive factors as well. The second key finding was that metacognitive approaches to instruction have been shown in the research to have a powerful impact on students ability to establish learning goals and monitor the progress (Donovan et al., 1999). Both of these key findings highlight that students come into classrooms with varying degrees of agency and a wide range of socially, environmentally, and experientially mitigated conceptions of themselves and their world.

The implications of these findings for learning and assessment are critical. Schools and classrooms are most effective when they are learner centered, meaning when educators are in constant study and awareness of the knowledge skills and attitudes that learners bring with them. Research has shown that students bring different mindsets into the classroom; some students believe that abilities or intelligence are fixed and unmalleable, while others believe that academic struggles allow practice and

development opportunities (Dweck & Leggett, 1988). The second key implication for the classroom is the importance of formative assessments; the focus of a formative assessment is to make students' thinking visible to both the teacher and the student. Formative assessment should be aimed at helping students see their progress over time and helping teachers identify or diagnose gaps in their learning for further development (Donovan et al., 1999). A critical piece in understanding the learning taking place in the classroom is assessment, and, as the state of California currently stands, we still have room to grow in the effort to integrate best practices of how students learn and grow into our assessment system.

Implications of Social Cognitive Theory and Human Agency on Assessment

Research has shown that learning and achievement in school is not simply about cognitive ability, IQ, academic skills, or the collection of knowledge. Rather, academic achievement is about noncognitive factors as well. Our assessments and school accountability measures should acknowledge this reality. There is a fundamental tension in our assessment system between assessing student learning through classroom-based instruments or large-scale instruments; they each serve their purposes, but their purposes and focuses are rarely aligned with one another. Ideal classroom-based assessments should serve formative purposes, and ideal large-scale assessments would likely serve summative purposes; it has been warned that unless careful balance is struck between these focuses, summative assessments will often overtake formative assessments (Black & Wiliam, 1998). This cannot be allowed to happen because in the process of learning and in formative assessments are the most genuine opportunities to see the noncognitive

factors of learning play prominently in academic achievement. Many researchers have argued that advances in cognitive science and psychometrics have demonstrated the power and potential of refashioning assessments to take advantage of those developments (Baker, 1997; Baxter, & Glaser, 1999; Linn, 1988; Messick, 1984; Mislevy, 1994; Nichols, 1994; Pellegrino, Jones, & Mitchell, 1999). Even with all of the effort to create new assessments through Common Core, we are still stuck with state assessments that do not address these research findings. Donovan et al. (1999) argued that assessments should find ways to balance the needs of instruction and assessment; it is not enough for them to simply tell the percentile rankings in summative fashion—they need to help inform targeted classroom instruction as well (Pellegrino et al., 2001). There are assessments out there that take advantage of the developments of psychometrics and cognitive science and provide both more targeted classroom instruction information and formative evaluation. One such assessment, the MAP Growth assessment, will be explored further in this research.

Achievement As It Relates to the California LCAP and LCFF Context

Over the last several years, California has moved away from the external accountability models propagated, mostly recently, under the NCLB era and extending into the Local Control Funding Formula (LCFF) and Local Control and Accountability Plan (LCAP) era. At the current time, schools and districts have been allowed to generate a more locally based context for educational reform instead of being purely focused on state assessment test scores. During this transition, Michael Fullan, well known for his work in system reforms, has been working with the state of California on a series of

feedback reports to monitor the progress made in California's quest to reform its education system. He argued that there are four main drivers that improve education efforts: capacity building, pedagogy, collaboration, and systemness (Fullan & Rincon-Gallardo, 2017). Overall, state leaders in education are excited and pleased with the new LCFF and LCAP funding and accountability structure because it establishes the ability to fund and focus on correct drivers for system reform; gains have been made in graduation rates, college readiness indicators have risen and suspension rates have dropped (Fullan & Rincon-Gallardo, 2017).

Within the new system, two of the LCFF priority areas are implementation of state academic standards and academic achievement. However, according to Fullan and Rincon-Gallardo (2017), "three in every five grade 11 students in California are ready or conditionally ready for college work in English language arts, and one in three are ready for college work in mathematics" (p. 8). Progress has been made, but there is still much room to grow in as many students are not prepared for higher level academics according to the most recent data. In 2018, the state released its new data dashboard, which is designed to provide basic data analysis of LCFF priorities for parents, schools, and districts. However, analysis from the CORE-PACE Research Partnership highlighted a major flaw in the dashboard that is really an indicator of broader deficiencies in the monitoring of academic achievement data. According to Polikoff (2019), the state has not fixed its approach to tracking student growth:

> The state has chosen a simplistic "change" measure by merely taking the difference between this year's scores and last year's scores on each outcome. This approach suffers from many problems, not the least of which is that it does not adjust for the fact that these are different students being compared to

one another (i.e., there are "cohort effects"). Especially for test scores, where there is a wealth of knowledge about the best ways to construct accountability system growth measures, there is no reason for the state to choose the approach it did. The state should choose a more appropriate growth measure, such as a two-step value-added model." (p. 5)

Simply put, we don't have clean methods to monitor and track academic growth within the basic frameworks of the state's LCFF priorities and the California dashboard. It is imperative that the system be able to objectively track and monitor student growth and find ways to help students use this growth data to increase their own self-efficacy. Fullan & Rincon-Gallardo (2017) argue that in order for California to build capacity (one of the four main correct drivers) the state must build capacity in "assessment with respect to defining, measuring, and using evidence for diagnostic, monitoring and action taking purposes" (p. 14). As it currently stands, state assessments provide only somewhat useful data; they currently define and measure but are difficult to diagnose and monitor because they are summative. Capacity to connect formative processes, such as benchmarks or other formative interim assessments, will be critical to bridging this gap.

Noncognitive Factors as They Relate to LCAP and LCFF

Noncognitive factors have been used to explain a relatively common phenomenon, "Numerous instances can be cited of people with high IQs who fail to achieve success in life because they lacked self-discipline and of people with low IQs who succeed by virtue of persistence, reliability and self-discipline" (Heckman & Rubinstein, 2001). This dichotomy has been explained, in various studies from a range of disciplines, as the interplay of academic outcomes and noncognitive factors (Bowles, 1976; Farkas, 2003; Heckman et al., 2006; Jencks, 1979; Lleras, 2008). In recent literature, a host of mental

constructs that are believed to have an impact on academic outcomes have risen to prominence in research. The dominant interpretation, as it currently stands, is that there are a host of behaviors, skills, attitudes and strategies that impact academic achievement but which might not be measured in any sort of meaningful and international way through classic academic assessments or instruments.

Educators know and experience a host of factors that impact a student's academic performance that do not appear to be cognitively based: attendance, responsibility, self-regulation, problem solving, beliefs about their own intelligence, persistence, and relationships with their peers and adults are some of the many factors that have been shown to make a difference in how students access and find success in school (Ames & Archer, 1988; Bandura, 1997; Bandura & Schunk, 1981; Conley, 2007; Farkas, 2003; Schunk & Hanson, 1985; Wentzel, 1998; Zimmerman, 1990). Thorough and rigorous research has demonstrated for many years the importance of mindset, essential skill development, and habits in shaping educational outcomes and achievement. However, one of the primary difficulties discussed in the literature of noncognitive factors and academic achievement is that there is still much to be understood about the degree to which any of these noncognitive factors are malleable and just how causal each of these factors is on academic achievement. Current research is trying to understand the interplay of these noncognitive factors with one another and upon academic achievement. Researchers are encouraged that investigation in this space could lead to positive breakthroughs because recent research has shown that even short-term interventions that target and address psycho-social beliefs can have an impact on academic performance

(Blackwell et al., 2007; Good et al., 2003; Oyserman et al., 2006; Walton & Cohen, 2007). How this research has shaped the policy context in K-12 education in California in recent years will be explored next.

The LCAP and LCFF era has also brought changes to the accountability system outside of cognitive academic measures. Today, schools and their districts are pushing towards a greater understanding or how noncognitive factors influence student wellbeing and academic achievement. However, K-12 schools are really just beginning the challenging work of integrating these noncognitive factors into their accountability systems and continuous improvement efforts. Nine key districts in California, collectively known as the CORE Districts, have led the effort to understand and integrate noncognitive factor development and integration into the California K-12 system. In 2013, the CORE Districts convened and selected four key noncognitive factors that they believed would best foster the development of mindsets, essential skill development, and habits (MESH). Their selection criteria were that MESH competencies must demonstrate in the research meaningfulness, measurability, and malleability (Kane & Mitchell, 1996). The four noncognitive factors chosen from these criteria were self-management, self-efficacy, growth mindset, and social awareness. The CORE districts and their partners have worked to develop a survey to measure and evaluate these four noncognitive factors within their districts and schools.

The CORE districts and their MESH work have been encouraged and driven by the longitudinal research that shows long term positive outcomes can be attributed to noncognitive factors. One study showed that noncognitive factors were just as likely to

27

predict college degree attainment as cognitive ability. In another study, kindergartners

with high social competence were shown to be more likely to graduate from high school

and college (Heckman et al., 2006). Two key policy recommendations to come out of the

MESH survey work conducted by the CORE districts were to begin systematically

measuring MESH competencies and to use those results for formative system

improvement efforts (Larocca & Krachman, 2016). The work of this dissertation is to

study two particular MESH competencies with the goal of understanding the impact of

those competencies on academic achievement.

Focus on Noncognitive Factors: Self-Efficacy and Academic Mindset

The literature has organized the noncognitive factors for academic achievement into

five broad categories: academic mindsets, social skills, academic perseverance, learning

strategies, and academic behaviors. Of these five factors, it has been hypothesized that

academic mindset is the noncognitive factor that impacts the other four factors and that it

is through this mindset lens that the other four factors derive much of their input

(Farrington et al., 2012). Figure 1 demonstrates this hypothesized model

Figure 1

Hypothesized Model of Five Noncognitive Factor Interaction

```
                        SOCIO-CULTURAL CONTEXT

                 SCHOOL AND CLASSROOM CONTEXT ──────────┐
                                                         │
   ┌────────────────────────────────────────────────┐   │
   │              ACADEMIC MINDSETS                   │   │
   └────────────────────────────────────────────────┘   │
        │              │              │            │     │
        ▼              ▼              ▼            ▼      ▼
   ┌─────────┐   ┌──────────────┐   ┌──────────────────┐
   │ SOCIAL  │   │  ACADEMIC    │◄──│    LEARNING      │
   │ SKILLS  │   │ PERSEVERANCE │   │   STRATEGIES     │
   └─────────┘   └──────────────┘   └──────────────────┘
        │               │
        │               ▼
        │        ┌──────────────┐
        └───────►│   ACADEMIC   │◄──
                 │   BEHAVIORS  │
                 └──────────────┘
                        │
                        ▼
                 ┌──────────────┐
                 │   ACADEMIC   │◄──
                 │ PERFORMANCE  │
                 └──────────────┘
```

STUDENT BACKGROUND CHARACTERISTICS
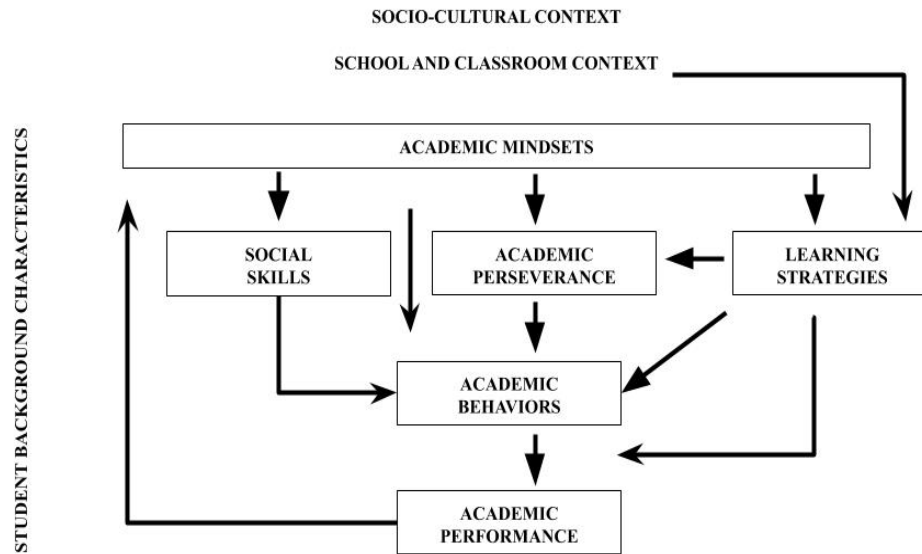
Note. From "Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance—A Critical Review" by C. A. Farrington et al., 2012, The University of Chicago Consortium on Chicago School Research.

Growth mindset and self-efficacy are two constructs within the academic mindset noncognitive factor structure that have been targeted by the MESH survey as the result of research that has shown their potency and relevancy to understanding student academic performance. The goal of this dissertation is to further explore the relationship between self-efficacy, growth mindset, and academic performance as measured by the MAP Assessment. In the hypothesized model provided, it is suggested that there is a feedback loop between academic mindsets, other noncognitive factors, and academic performance. In K-12 education, grades and standardized testing scores are the two primary measures

of academic achievement. This dissertation explored how academic mindset could be impacted by the MAP Growth assessment, an assessment that mirrors some features of state testing but does so in a way that is more formative in nature and could have a more useful impact on teaching, learning, and noncognitive factor development.

Growth Mindset

Dweck (2008) compared individuals with fixed mindsets—those who believed that traits were given and that talent determines success—to individuals with growth mindsets—those who believe that basic abilities can be refined through dedication and hard work. The findings matched those of other researchers: that those students who believe that effort and hard work matter are more interested in building capacity, more persistent, and more likely to display other noncognitive factors tied to achievement (Cury et al., 2006; Dweck & Leggett, 1988).

Development of Academic Mindset & K-12 Achievement. Many longitudinal studies have demonstrated the connection between growth mindset and students' motivation, grades, and higher test scores (Transforming Education, 2016). Growth mindset has been established as an influential factor on achievement and numerous follow up studies have been conducted to better understand how mindset can be cultivated and developed in the classroom. Research has shown that that adaptive motivational patterns, or growth mindset, come from academic processes that incorporate challenge and even failure while at the same time directly supporting the motivational context of the student (Dweck, 1986). In subsequent research, it was found that when those challenging tasks were focused on skill acquisition rather than evaluation, students,

regardless of ability, were more willing to take on more challenging tasks in order to master the skills of that particular domain (Dweck & Leggett, 1988). The implications of these two studies together is that in a classroom context that is challenging, safe to struggle within, and supported adequately, students can develop strong academic mindsets that will encourage them in academically challenging pursuits, regardless of their initial ability level. A synthesis of research from several studies found that when teachers make challenging tasks accessible to all students, providing support to help students achieve success, students are more likely to rise to the challenge (Dweck et al., 2014). Growth mindset is hampered when goals are not geared toward skill development. In a series of studies, evidence was found that non-evaluative learning goals could have a strong impact on student motivation and performance (Grant & Dweck, 2003). In this context, failures were seen as part of the growth process rather than as indications of inability.

Research has demonstrated the potential for maladaptive strategies and reinforcers that can prevent growth mindset. In one study, Dweck and Reppucci (1973) were able to predict and correlate 5th graders responses to why they were unable to solve difficult math problems based on a questionnaire run before the experiment. The "Intellectual Achievement Responsibility" questionnaire was designed to detect whether a student would blame others or themselves for their inability to solve the problems. The study found that the students who blamed themselves were also more likely to be able to solve the difficult problems, while those who were most likely to blame others were more likely to not be able to solve the problem (Dweck & Reppucci, 1973). In a follow up

study by Dweck et al. (1978), 5th graders were systematically observed to evaluate the types of feedback they received from their teachers. The feedback pattern suggested that boys were far more likely to receive negative feedback based on rules, effort, or their inability to follow class norms. During the second phase of the same study, the students' attitudes towards their feedback were studied. What the researchers found was that boys were far more likely to blame their teacher for their academic performance, while girls were more likely to blame themselves. The researchers suggested this was due to the feedback originally provided by the teachers; since the boys' feedback was more often based on not following the norms and rules and not their actual academic performance, they were less likely to have their mindset towards their own ability impacted. The girls, on the other hand, were more likely to receive feedback about their academic performance, because they were more likely to be following the class norms; consequently, that feedback had a greater potential to lead the girls to question their academic ability (Dweck et al., 1978). Research into the space of academic mindsets and growth mindset has led to the formation of two categories of students: those who are performance oriented and those who are mastery oriented. Students who are performance oriented are more likely to evaluate themselves poorly in ability or predict failure for themselves, while mastery-oriented students are more likely to self-motivate, focus on their potential, and predict that success will eventually come (Smiley & Dweck, 1994).

Growth Mindset & Middle School Math Achievement. The transition from elementary to middle school can be a difficult one for students. In middle school classes become more challenging, expectations increase, the workload increases, grading is more

challenging, and the structure of instruction is often less personalized than that of elementary school (Midgley et al., 1995). West and Schwerdt (2012) studied this transition within the context of two different types of schools: K-8 schools and 6-8 middle schools. In the K-8 schools, it was found that achievement level drops during the transition from 5th to 6th grade within the same school were lower than the achievement level drops of students transitioning from an elementary to middle school (West & Schwerdt, 2012). It has been suggested that, in addition to the rising expectations and other environmental factors mentioned previously, middle school achievement drops in part because of the changing goal orientation between elementary and middle school; Midgley et al. (1995) found that elementary teachers were more likely to focus on task or learning goals while middle school teachers were more likely to focus on performance goals. This change can have a negative impact on the mindset of a student. When students focus on performance goals, they will often evaluate their own success in relation to their peers and potentially judge themselves to be incapable of achievement if the task is difficult and they are unable to perform. Research has shown that growth mindset is malleable and with intervention it can be positively impacted.

In a widely-cited dual study of growth mindset in the context of middle school mathematics, Blackwell et al. (2007) studied whether achievement could be predicted by a student's implicit theory of intelligence. The first study followed students as they transitioned to middle school and moved through middle school; students completed scales designed to assess their growth mindset and their achievement scores on standardized tests were tracked. Blackwell et al. (2007) was able to demonstrate that

incremental theory of intelligence (growth mindset) was positively associated with beliefs

about the importance of effort, what kind of learning goals the students had, positive

work habits, and fewer feelings of helplessness. In their second study, Blackwell et al.

(2007) wanted to take what was found in the first study and discover whether a positive

theory of intelligence and the correlating positive motivational patterns could make an

impact on achievement. In this study, a group of students was evaluated, and each was

identified as having either a growth or fixed mindset, then they were randomly assigned

to either the control or experimental group. The experimental group spent eight 25-

minute advisory sessions discussing how the brain is like a muscle and can be trained just

like other muscles. At the end of the study, students who were given the advisory

sessions had GPAs an average of 0.30 points higher than students who did not;

additionally, it was found that students could change their mindsets in just a few months

(Blackwell et al., 2007).

There has been some indication about the power that short-term interventions have

had on helping students to develop growth mindset. The research conducted by Blackwell

et. al (2007) was built upon previous research in which a similar intervention was

conducted amongst college students. In that previous study, researchers working with

college students showed a video explaining that academic setbacks are normal. At the end

of the study, students who saw the video had GPAs an average of 0.27 points higher than

students who did not see the video (Wilson & Linville, 1985). In both cases, relatively

low-impact interventions providing basic information on growth mindset had significant,

measurable impacts. In a widely-cited study synthesizing the findings of many studies on

academic-mindset, Yeager and Walton (2011) found that:

> Several rigorous, randomized field experiments have shown that seemingly
> "small" social-psychological interventions—typically brief exercises that do
> not teach academic content but instead target students' thoughts, feelings, and
> beliefs in and about school—have had striking effects on educational
> achievement even over months and years." (p. 2)

In a follow up study, a one-hour online growth mindset intervention, designed to be cost

effective, scalable, and widely deployable, was tested nationwide. The study asked

students to complete two 25-minute online sessions in which they read and listened to

scientific materials about how the brain works and could grow. The treatment session

also encouraged students to reflect on how to apply this to something they might want to

grow their brain in pursuit of and how to put these beliefs into practice (Yeager et al.,

2019). The goals of the study were twofold: to determine if such a short-term intervention

would have a positive impact, as previous interventions shown to have an impact took

longer to implement, as well as to determine in which types of schools the intervention

would have the greatest impact. Yeager et al. (2019) found that the growth mindset

intervention:

> reduced the prevalence of fixed mindset beliefs relative to the control
> condition, reported at the end of the second treatment session, unstandardized
> $B = -0.38$ (95% confidence interval $= -0.31, -0.46$), standard error of the
> regression coefficient (s.e.) $= 0.04$, $n = 5{,}650$ students, $k = 65$ schools, $t =$
> -10.14, $P < 0.001$, standardized mean difference effect size of 0.33. (p. 366)

Furthermore, their academic performance improved as well. Yeager et al. (2019) found:

> lower-achieving adolescents earned higher GPAs in core classes at the end of
> the ninth grade when assigned to the growth mindset intervention, $B = 0.10$
> grade points (95% confidence interval $= 0.04, 0.16$), s.e. $= 0.03$, $n = 6{,}320$, $k =$

65, t = 3.51, P = 0.001, standardized mean difference effect size of 0.11, relative to comparable students in the control condition. (p. 366)

According to Yeager et al. (2019), the effects were most pronounced in low and medium achieving schools that had a school culture that embraced growth mindset and had student populations that provided support to their peers for academic risk taking. It could be very tempting for school districts and sites to view this research and see the relatively quick and easy opportunities for deployment. However, it has been warned that viewing these interventions as silver bullets without considering the specific contexts of the areas in which they were deployed would be a significant mistake, and that rigorous and careful construction of these interventions is necessary (Yeager & Walton, 2011; Yeager et al., 2019).

Self-Efficacy

Self-efficacy has found a wide range of applications in many fields of study. In my research for this work a search for "self-efficacy" in my university's library database generated 374,442 results and a Google search of the same generated 119,000,000 results. The concept of self-efficacy originated from Albert Bandura, a foundational theorist who is known for two theoretical frameworks: self-efficacy, and the larger theoretical framework of social cognitive theory. He defined self-efficacy as "beliefs in one's capabilities to reorganize and execute the courses of action required to produce given attainments" (Bandura, 2012, p. 3). Success, or lack of success, in executing those courses of action would further impact the attitudes, beliefs, and future decision making of that individual. Bandura (1977) believed that individuals generated powerful self-perceptions from their experiences and ability to achieve success in tasks. While self-

efficacy was theorized to be a powerful force, this power was qualified with some

limitations. Bandura (1977) states:

> Expectation alone will not produce desired performance if the component
> capabilities are lacking...given appropriate skills and adequate incentives,
> however, efficacy expectations are a major determinant of people's choice of
> activities, how much effort they will expend, and of how long they will
> sustain effort in dealing with stressful situations." (p. 194)

Self-efficacy fit into a wider theoretical framework of social cognitive theory. Bandura

(1986) posited that human achievement is derived from the interaction of three forces:

one's own behaviors, personal factors or characteristics, and environmental conditions.

Within the theoretical framework of social cognitive theory, the central concept of

reciprocal determinism states that these three forces dynamically and reciprocally interact

with each other to influence an individual's goals, values, and self-efficacy beliefs. This

interaction has huge potential significance in education. Social cognitive theory suggests

that economics, personal ability level, or a host of other factors do not single handedly

pre-determine academic outcomes but rather these factors dynamically interplay with

each other to form the self-efficacy of any given individual. This formed self-efficacy

will then loop back to one's own behaviors and actions, thus shaping their academic

outcomes.

According to Bandura (1986), four main factors influence self-efficacy: mastery

experiences, vicarious experiences, social persuasion, and physiological states. Mastery

experiences are those experiences in which an individual has experience completing a

difficult task and gains positive self-beliefs from that enterprise. Bandura (2012)

described these mastery experiences as "the most influential source of efficacy

information because they provide the most authentic evidence of whether one can muster whatever it takes to succeed" (p. 80). Vicarious experiences are those in which an individual gains a rise to their self-efficacy by watching or experiencing modeled attainments of success. Vicarious experiences are complementary to mastery experiences because the social nature of human existence often allows us to compare our mastery in relation to others. Social persuasion is the process in which an individual can be encouraged to persevere (or not) through difficult situations with the encouragement of a role model or influential person; these influences can be both positive and negative. The last major influence of self-efficacy is physiological status: success or failure can create such degrees of euphoria or stress that it can impact the bodily functioning of an individual. Acute levels of stress or anxiety derived from failure can impact an individual's self-efficacy.

Both self-efficacy and self-concept have received attention in research. Sometimes, they are used interchangeably when they are, in fact, two different concepts. Self-concept can be defined as an individual's belief about their abilities, often through a more general evaluation; self-efficacy can be defined as the beliefs an individual has about their ability to perform a specific task or activity effectively (Pajares & Schunk, 2002). Self-efficacy has been shown in the research to be a stronger indicator of potential academic achievement then self-concept. When tested against each other in middle school, high school, and college settings, self-efficacy has been found to be the more predictive factor for academic achievement than self-concept (Pajares, 1996; Pajares & Graham, 1999;

Pajares & Miller, 1994; Pietsch et al., 2003). This suggests that the self-efficacy has some fundamental and underlying power, if, regardless of age level, it has an impact.

Development of Self-Efficacy & K-12 Achievement. It is not enough to merely be self-efficacious with no competencies or to have competencies with no self-efficacy; rather, these two have to be cultivated in tandem. As Bandura (2012) says, "Children have to learn to face displeasing realities about gaps in their knowledge and competencies" (p. 176). Students with low self-efficacy may simply believe they also have low competencies, while those with low competencies may form low self-efficacy as a result. Much research has been done on the successful development of self-efficacy. Easy tasks cannot increase self-efficacy; rather, the task has to be sufficiently difficult and appropriately supported so that the student could complete the task and increase their agency in the process (Bandura, 1977). Over time the supports could be removed and the student could still complete the task, thus increasing their self-efficacy. In one widely-cited study, Bouffard-Bouchard (1990) demonstrated that self-efficacy had a motivational component that correlated directly with the persistence shown by students on difficult tasks and their willingness to engage in overt inputs to produce a desired outcome. This study reveals the positive feedback loop that self-efficacy work creates: increased self-efficacy results in an increased drive to accomplish difficult tasks, and with the accomplishment of difficult tasks comes greater self-efficacy.

A broad range of studies have demonstrated the power and significance of self-efficacy in shaping students' academic and social outcomes. The proliferation of research on this topic has demonstrated that this may be one of the most promising areas on which

to focus school improvement efforts. In one meta-analysis conducted on 36 studies between 1977 and 1988, it was found that 14% of the variance in academic achievement could be attributed to self-efficacy, and that mathematics studies tended to have higher effect ratings than those in other academic areas (Multon et al., 1991). In another meta-analysis of three causal models for self-efficacy, skill development (student achievement causes self-efficacy), self-enhancement (student self-concept drives achievement), and the reciprocal model (they both drive each other simultaneously) were tested against each other (Valentine et al., 2004). The reciprocal model proved to be the most effective and provides insight into the components for a necessary intervention: it should address achievement and self-concept at the same time in order to have the greatest impact. In a study of junior and senior high age students, it was found that students with high ratings of self-efficacy were better able to monitor their own performance and self-regulate than those with low self-efficacy (Bouffard-Bouchard et al., 1991; Valentine et al., 2004). Academic motivation and persistence were also shown to be directly correlated to strong personal senses of self-efficacy, both before and after learning tasks took place. In addition, self-efficacy was found to be a better predictor of persistence than other variables tested (gender, grade level, and prior experience) (Pajares & Miller, 1994; Schunk, 1995). Self-efficacy can be trained, nurtured, and supported within the proper educational context, and this can have positive outcomes on preventing math anxiety and improving problem solving performance (Pajares & Kranzler, 1995).

Goal setting is an important component of developing one's self-efficacy (Bandura & Cervone, 1983; Schunk, 1983, 1996). In a study of 4th graders, those who set learning

goals rather than mere performance goals were more likely to improve their motivation, self-regulation, and self-efficacy (Schunk, 1996). It is challenging to set goals for learning without a fairly accurate awareness of where one's current achievement level lies. Schunk (1983) demonstrated in his study of elementary students the power of pairing goal setting with incentives in order to improve the self-efficacy of those students. In the study, students who had incentives for growth assigned to specific sub skills accelerated their learning and had marked improvement in their self-efficacy ratings. In order to target subskill growth and target incentives, a teacher would need specific diagnostic knowledge of a student's current abilities. Bandura and Cervone (1983) demonstrated another set of factors that contributed to improve self-efficacy in their study of college students on a non-math related task. In their study, they found that setting personal goals for achievement combined with timely feedback helped improve students' self-efficacy ratings. For students to access diagnostic sub-skill level information in order to inform their goals, it would help if more emphasis was placed on formative methods for assessment. It has been argued that in many traditional mathematics classrooms, students are rarely taught to understand the broader landscape of what they are learning by their teachers; however, when students learn to set goals based on formative feedback, they have higher levels of achievement and self-efficacy (Boaler, 2016). Bandura (2012) also argued that "Educational practices should be gauged not only by the skills and knowledge they impact for present use but also by what they do to children's beliefs about their capabilities" (p. 176). Students that are given opportunities to goal set and be

appropriately supported while accomplishing difficult tasks are more likely to increase their self-efficacy.

Self-Efficacy and Middle School Math Achievement. Self-efficacy and math achievement has been studied at the middle school level. Pajares and Graham (1999) wanted to determine the significance of self-efficacy beliefs on math achievement in middle school by controlling other motivation and achievement variables. In their study,12 variables were evaluated for their correlation to math achievement. The study also wanted to determine the level of change in self-efficacy that took place throughout the school year. This study found self-efficacy to be the only predictor of performance that at the beginning and end of the year when all other variables were controlled; students with declining performances generally had corresponding drops in self-efficacy and the reverse was true as well (Pajares & Graham, 1999). This study was particularly noteworthy because of the wide range of other variables studied, including engagement, gender, self-concept, and anxiety.

In another study, self-efficacy, intrinsic motivation, and participation in an after-school program were examined in a group of middle school Latinx students to determine the impact of those factors on academic achievement (Niehaus et al., 2012). The researchers wanted to determine whether self-efficacy could lead to differing academic outcomes and whether the after-school program provided could impact the levels of self-efficacy. The researchers ran a regression analysis of a self-efficacy assessment, an intrinsic motivation assessment, GPA data, school attendance data, and reading and math achievement data (from standardized test scores) (Niehaus et al., 2012). The researchers

found that self-efficacy was a significant predictor of math achievement, but self-efficacy

levels and growth amongst individual students varied across the sample who received the

intervention and the researchers acknowledged they could not explain this with their

study. This study aligns with previous research that shows self-efficacy is an important

factor in academic achievement.

Challenges To Academic Mindset

Meta-analysis suggests there may be stronger areas of academic intervention or

development that could have a larger impact on student achievement. The impact of

growth mindsets over fixed academic mindsets had an effect size of 0.16 (which is a

relatively low effect size) according to three meta-analysis studies including 237 studies

and 451,287 students (Hattie, 2018b). Furthermore, two other meta-analyses conducted

cast further doubt on the potential impact of academic growth mindset interventions on

student's academic achievement and the relationship between mindset and academic

achievement (Sisk et al., 2018)**.** In the first meta-analysis, it was found that only 100 of

273 total studies included in the meta-analysis had a statistically significant correlation

between growth mindset and academic achievement, 16 studies had a negative

correlation, and 157 studies were not significantly different from zero (Sisk et al., 2018)**.**

In the second meta-analysis, 37 of the 43 studies regarding growth mindset interventions

and academic achievement were not significantly different than zero, 1 study was

negative (suggesting that students were worse off after the intervention), and the five

remaining studies were significantly different than zero and positive (Sisk et al., 2018).

These findings would suggest that attempting to change mindsets alone, without

addressing any of the other known factors to impact student academic achievement, would in many cases not have much impact. However, Sisk et al., (2018) found that adolescents, typical students and students facing situational issues like school transitions received little benefit but there were some findings to suggest that high risk students and socio-economically disadvantaged students could find benefit. Furthermore, Sisk et al., (2018) acknowledge that academic mindset interventions are relatively cheap and easy to develop and could still create a net benefit when deployed with at risk students who are more likely to benefit from the intervention and further suggested that mindset interventions may need to be combined with other interventions to have a more definitive impact.

Challenges To Self-Efficacy

One research study on high school students in an online school setting sought to study the engagement and motivation of high and low performing students; researchers investigated how motivation, regulation and engagement shifted in these students throughout the course (Kim et al., 2015). Self-efficacy was one of the key variables used to measure the motivation construct in the study and the researchers found that high performers tended to have higher starting values for self-efficacy than low performers. Furthermore, low performers tended to have diminished self-efficacy throughout the semester (Kim et al., 2015). Two other constructs were measured in the study: regulation and engagement. The study's findings on regulation and the potential mediating impact on self-efficacy are critical. The regulation construct was divided into three variables: measured metacognitive regulation, the management of cognition in learning activities

and effort regulation, and the management of effort in learning activities when faced with difficulties (Kim et al., 2015). The study found that both high and low performers' effort regulation drifted over the semester and that the metacognitive regulation of both high and low performers decreased throughout the semester. The researchers suggested that this could help demonstrate that achievement not only depends on cognitive regulation but on how students control their emotions and motivation as well (Pintrich, 1999). All of this would further suggest that there could be distortions in measuring self-efficacy when students are in a state of poor self-regulation.

Perhaps combining academic mindset interventions with self-efficacy could have that more definitive impact. A recently updated meta-analysis on factors related to student achievementfound that self-efficacy had an effect of 0.71 according to eight meta-analysis studies involving 418 individual studies and 261,761 students (Hattie, 2018c). These numbers suggest that self-efficacy has a better-than-moderate impact on student achievement. Not surprisingly, almost all of Hattie's top ten items include definitive strategies for instructional change that would fundamentally change what many students experience in schools. However, the number one factor for student achievement is collective teacher efficacy (Hattie, 2018a). Collective efficacy is the joint belief regarding the total sum of abilities for a group to plan and carry out a desired outcome (Bandura, 1997). This would suggest that between self-efficacy having a better-than-moderate impact and collective teacher efficacy having the highest impact on student learning, efficacy is a very important influence on student academic achievement. It could be suggested that one reason for the difference in impact between student self-efficacy and

teacher collective efficacy is the connecting factors that may be influenced by the corresponding ratings of efficacy. With regard to student self-efficacy, this could certainly connect to and inspire corresponding positive impacts into other noncognitive factors and achievement. Collective teacher efficacy could have this same impact in addition to making it more likely that teachers would be willing to engage in the hard work and development of other high-leverage teaching strategies that have been shown to yield positive student outcomes.

History of High Stakes Testing Dangers and Opposition

Standardized testing and its role in public education is a topic that elicits strong opinions from educators and lay people alike. The Cold War space race and President Reagan's *A Nation at Risk* report helped ignite a fear that American schools were falling behind (National Commission on Excellence in Education, 1983). This led to a determination that students needed to be tested more and schools held accountable for the learning taking place. The extent of this fear was clearly demonstrated by a 2002 Gallup Poll in which 57% of Americans approved of the No Child Left Behind Act and 68% of Americans were in favor of nationalized standards (Wirt & Wirt, 2009). Today, standardized testing exists in a precarious and contradictory position in educational discourse. On one hand, it offers the opportunity to "objectively" verify growth and the success of our turnaround efforts; on the other hand, it has been demonstrably proven to have adverse effects on students and is heavily susceptible to corruption. Testing has risen in prominence in the educational system and many political concerns have arisen as a result. Critics of testing point to a host of concerns: invalid and biased testing

instruments; the dangerous rise of an economic and corporate testing culture; the harmful

effects of standardized tests on minoritized youth, students with special needs,

marginalized, and low SES students; corruption caused by cheating; and the punishments

districts and states issue to students and teachers as a result of test scores.

How Standardized Testing Supports Economic & Corporate Culture

Mary Anne Raywid focused particularly on the dangers that economic and corporate

thinking pose for education. She wrote:

> The economic/business analogy seems to have shaped and propelled the drive
> for accountability in education during the last decade. Since there are no
> profits to serve as indications of whether or not schools are doing a good job,
> test scores have been assigned that function instead. The insistence on
> quantitative measures of school effectiveness has reduced educational
> outcomes to testable products and de-emphasized the role of the school in
> other areas, such as preparing young people for civic participation,
> encouraging their personal development, and helping them master higher level
> intellectual skills. (Berliner & Biddle, 1996a, p. 194).

For the corporate- or economically-minded advocates of standardized testing, high stakes

accountability provides an opportunity to incentivize, punish, and reward; this theory has

provided the framework for many initiatives to motivate students and teachers (Herman

& Haertel, 2005). Competition is one of the chief economic forces at work in driving

efficient and productive business; it is often assumed that testing will trigger competition

and therefore more efficient and productive educational systems. But in the context of

education, competition's dangers have been demonstrated. One of the effects of zero-sum

competition has been the incentive to protect competitive advantage by teaching only the

students who will bring in the most money or prestige (Amrein-Beardsley & Collins,

2012). Competition also often works against cooperation and collaboration. Nichols &

Berliner (2007) highlight one of the greatest obstacles to applying economic theories of motivation, competition, and incentives to education, namely, that we do not control the inputs to educational "products" in nearly the same way that inputs to economic products can be controlled by their producers (p. 20).

How Testing Marginalizes Minoritized, Low SES, English Learners & SPED Students

Proponents of standardized testing will argue that they are trying to assess the learning taking place between the student and teacher in the classroom. One common argument of standardized testing proponents is that "testing gives the teacher important diagnostic information about what each child is learning in relation to what is taught" (Gross, 2013). However, much research has shown that standardized testing scores and the quality of teaching is not clearly correlated. Research has shown the importance of out-of-school factors in determining the variation of student achievement between different schools and communities. According to Berliner (2013):

> Out-of-school variables account for about 60% of the variance that can be accounted for in student achievement. In aggregate, such factors as family income; the neighborhood's sense of collective efficacy, violence rate, and average income; medical and dental care available and used; level of food insecurity; number of moves a family makes over the course of a child's school years; whether one parent or two parents are raising the child; provision of high-quality early education in the neighborhood; language spoken at home; and so forth, all substantially affect school achievement." (p. 5)

Opponents of standardized testing, particularly the high stakes forms in which sanctions or important student outcomes are triggered, are critical. The students and schools with the highest likelihood to struggle are also the most likely to be punished based on their "performance," while the deeper societal factors contributing to much of the performance

48

gap are outside of the control of the school itself (Berliner, 2013). On multiple occasions, educators across the country have been documented as administratively withdrawing students deemed "too far gone" to help, focusing on bubble students, and administering "regular exams" for special education students (Nichols & Berliner, 2007, p. 62-65). The challenges faced by English Learners have been well documented. In Texas and North Carolina, EL students were 40 to 60% less successful on the standardized exams then their white middle-class cohort mates (Nichols & Berliner, 2007). When standardized testing had its highest stakes, during the era of NCLB, some educators and states saw these stakes and reacted in ways that were not to the benefit of students.

How Standardized Testing Is Used to Unfairly Punish Teachers

Standardized testing can be "high stakes" for both students and teachers. Some teachers have been subjected to Value Added Modeling (VAM), in which standardized test scores are used to evaluate their performance. One of the major problems in assessing teacher performance based on standardized test scores is, as a variety of research has shown, that only 1% - 20% of the variance can actually be attributed to teacher performance (American Statistical Association, 2014; Fantuzzo, LeBoeuf & Rouse, 2014; and Haertel, 2013, as cited in Berliner, 2018, p. 7-8). Teachers are naturally resistant to being held personally financially accountable for test results, or being told they are ineffective teachers, when they know what the research verifies: that their own teaching has only a relatively minor impact on student standardized test performance. VAM is additionally problematic because the stability of teacher performance from year to year is variable. The Bill and Melinda Gates Foundation (2012), as cited in Berliner (2014), tried

to correlate teacher performance and testing achievement through an in-depth observational study. Their data demonstrated that outside-of-school factors made it difficult to reliably predict student achievement on the tests based on teacher performance alone Formatting.... Teachers pinned between VAM evaluations and working with some of the nation's disadvantaged students are put in an unenviable position; some may even feel compelled to seek unethical means not to fall behind.

How Standardized Testing Is Gamed and Cheated

When standardized testing becomes high stakes testing, it is particularly susceptible to corruption; this is due to a phenomenon known as Campbell's Law. Campbell's Law states that "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor" (Campbell, 1975, as cited in Nichols & Berliner, 2007, p. 26-27). In education, this "law" has demonstrated itself in a wide range of "low level" cheating incidents as well as several cheating scandals that have reached national level discussion. Educators across the country have been documented providing testing materials and questions to students before the test, providing unauthorized help or cues during the test, and even "scrubbing" scores after the test (Nichols & Berliner, 2007, p. 46-47). It is hard to gauge exact numbers, but a national survey and a study done in the Chicago Public School system suggests that approximately 5-10% of educators may be helping their students cheat on standardized tests (Nichols & Berliner, 2007, p. 46-47). Sometimes these cheating scandals rise above individual actors operating at the classroom and site level and become systemic

conspiracies. In well-documented testing scandals in Atlanta, GA; Washington, D.C.; Denver, CO; and Houston, TX, the cheating involved tacit district support and involvement as well (Nichols & Berliner, 2007). When an educator's job may be on the line, bonuses or prestige may be gained or lost, or student graduations and promotions are at stake because of a single test, Campbell's Law tells us there is an extreme risk of distortion.

NWEA's MAP Growth Assessment and State Common Core Assessment

As discussed previously in this review, California has a problem: tracking academic growth at the state or individual site level is imprecise. At the state level, cohorts are being compared to other cohorts only on a year-to-year basis. For a variety of reasons, these types of comparisons are not useful to track growth. At the individual site level, assessments taken at the end of a year provide no useful or actionable instructional data for the cohort of students that actually took the exam. In a cross-grade level study teachers were interviewed about state assessments. Four common themes arose: there is inadequate diagnostic information from the state tests, test scores come too late to inform future instruction, there is no baseline data or ability to track growth within students, and the tests are designed for average students only and do not properly account for high or low achievers (Yeh, 2006). The MAP Growth assessment may provide a way to bridge this gap. According to Yeh (2006), those same respondents, when interviewed about the MAP Growth assessment, reported that the adaptive tests provided more diagnostic tools, results were available more quickly, the adaptive tests demonstrated progress, and the tests were able to applicable to all ranges of students. According to NWEA, MAP Growth

assessments are accurate predictors of proficiency on California's state assessments (Northwest Evaluation Association [NWEA], 2017). If the MAP Assessments can predict levels of success on the California SBAC, and the MAP Growth assessments have the added feature of being able to provide growth data throughout the year as well, it provides educators the opportunity to use data generated by norm-referenced standardized tests to inform classroom practice.

Development of NWEA MAP Growth Assessment

Starting in the 1970's, Northwest Evaluation Association (NWEA) was formed. Their mission was to create an assessment that was adaptive and was based on a Rasch item response theory (IRT) model. The earliest iteration of what would eventually become the MAP Growth assessment included the Achievement Level Test, which measured normal progression, falling behind, or accelerating progression in math and English (NWEA, 2019). In the 1980s, they developed their first adaptive assessments, but due to limitations in technology, these adaptive tests were used only in a limited fashion. The advent of widespread technological deployment in schools in the early 2000s allowed the MAP Growth assessment to be more widely deployed and utilized (NWEA, 2019). The MAP Growth assessment is a hybrid assessment; it contains elements of a summative standardized test and elements of a formative interim assessment.

Design of NWEA MAP Growth Assessment

NWEA built their MAP Growth assessment around a few important design principles that are worth exploring and understanding. The first key design element of the MAP Growth assessment was IRT modeling. The advantage of IRT modeling is that it allows

for item banks to be developed that have defined characteristics that can be applied not only to the testers but to future testing populations as well (Rasch, 1980). This allows testers to be normed against each other. The second key design element that was essential for the assessment was adaptability. Adaptive tests draw future questions for a student based on their performance on previous questions (Weiss, 1974). The advantage of adaptive testing over fixed-item testing is that it allows the low- and high-end students to be more accurately assessed, instead of just the middle grouping of students that are typically served by a fixed-item assessment. Lastly, NWEA closely follows the Universal Design for Learning (UDL) principles, which recognize that the students tested come from a wide variety of backgrounds and are designed to ensure test validity and reliability (Thompson et al., 2002; Weiss, 1974). It is through these three design principles that NWEA hopes to accomplish one of its six core guiding principles for the purpose of the assessment: "Provide information about a student's change in achievement level from one test occasion to another, as well as the student's current achievement level. A single test result is only a snapshot of student achievement. Multiple snapshots are needed to gauge a student's growth over time" (NWEA, 2019, p. 8). NWEA's technical and thorough analysis of their MAP Growth assessment suggests that by using these design elements they have created a valid and reliable assessment.

Validity of NWEA MAP Growth Assessment

Five important criteria exist to assess the validity of an instrument: content evaluation, response processes, internal structure, relation to external variables, and consequences (American Educational Research Association [AERA] et al., 2014).

Questions on the NWEA map assessment are carefully designed using Webb's three levels of depth of knowledge (DOK) to ensure that all range of students can access the assessment (Webb, 1999). Additionally, the MAP Growth assessment has been independently studied by an outside auditor. This review randomly sampled 20% of the questions used in the MAP growth assessment; in total, 1,563 Reading items, 1,134 Language items, and 1,702 Mathematics items were studied. The study found that, on average, 97.4% of the items were aligned to the Common Core across all grades and content areas (Egan & Davidson, 2017, as cited in NWEA, 2019).

The internal structure of the assessment has been shown to be sound through the differential item functioning (DIF) detection processes used by the MAP Growth assessment. In the most recent technical report, 500 mathematics items from the pool received a DIF analysis to ensure internal structure. That DIF found those math items, when examined by the race and gender of the students, had no less than 83 percent of the items received anything lower than an "A" rating by ETS standards. An "A" standard represents negligible levels of DIF (NWEA, 2019; Zwick, 2012).

The relationship between NWEA and other external variables has been established as well. According to the National Center on Response to Intervention (NCRTI) correlations must exceed 0.70 to demonstrate concurrent validity (American Institute for Research, 2016). At the middle school level concurrent validity ratings are in the .84-.83 range and the classification accuracy range is in the 0.84-0.82 range (NWEA, 2019, p. 94). Similarly, the California SBAC assessment for 8th grade has a concurrent validity rating

of 0.85 with the NWEA's MAP Growth RIT score, suggesting a higher degree of external validity between the two assessments (NWEA, 2019).

An area of weakness in the validation of the MAP Growth Assessment are that NWEA has focused much of its effort on studying response processes along test disengagement lines. Much of their research and efforts in developing the MAP Growth assessment in this area have focused on preventing test disengagement through technological cues for the proctor and students. NWEA claims that disengaged test taking can occur in low stakes tests like their MAP Growth assessment but that students do not usually disengage for the entire assessment (Wise & Kong, 2005; Wolf et al., 1995). Also, NWEA has not thoroughly reviewed the consequences of their instrument; this is likely because, for their own stated purposes of the assessment, namely, low stakes diagnostic and formative assessment, there should be little negative risk.

Reliability of NWEA MAP Growth Assessment

Criteria also exists to assess the reliability of an instrument. Three major categories of reliability can be assessed; they are usually represented as coefficients and include alternate forms, test retest, and internal consistency (AERA et al., 2014). A fourth criteria for evaluating the reliability of an instrument also exists—assessment of the rater—but as the MAP Growth assessment is machine scored, this will not be explored further here.

The NWEA MAP Growth assessment has been shown to have a high degree of internal consistency and has acceptable levels of test-retest reliability. Due to the adaptive nature of the MAP Growth assessment, traditional forms of reliability validation are difficult. However, NWEA uses a coefficient alternative they find just as reliable that

uses measurement error at different achievement levels to index and help demonstrate the reliability of the assessment (Samejima, 1977, 1994). The internal consistency of the MAP Growth assessment at the middle school math level ranges from 0.905-0.919, a very high degree of marginal reliability (NWEA, 2019). Similarly, measuring test-retest reliability is not possible with traditional methods of measurement. NWEA uses a hybrid form of test-retest that resembles an alternate forms measurement (Crocker & Algina, 1986). The hybrid test-retest and alternate forms reliability measures generally rate very well for the math MAP Growth assessment, with a range of 0.916-0.915 at the middle school level (NWEA, 2019).

The validation and reliability measurement conducted by NWEA suggests, as earlier claimed, that the MAP Growth assessment is strongly aligned with California's state assessment. As such, there is great potential in utilizing this assessment for the purpose of this study; not only is it aligned with the state's chief academic instrument, but it also includes additional features that may make it more useful that other standardized tests in the context of the classroom and instruction.

It has been argued that assessment must be restructured to reinforce and strengthen growth mindset practices. The testing regime has done much to prevent this restructuring from taking place; instead, the focus on grades and scores has created mathematics students who develop fixed mindsets (Boaler, 2016). Studies have shown that students are quick to identify themselves their math ability by their grades or test scores (Kohn, 2011; McDermott, 1993). Furthermore, studies have shown that grades can actually have a negative impact on students' achievement and that students performed better if they

received formative/diagnostic feedback alone (Butler, 1987, 1988; Elawar & Corno, 1985; Pulfrey et al., 2011). The MAP Growth assessment's simultaneous focus on achievement and growth, if properly framed with students, may help to shift the conversation towards more formative conversations of student achievement and provide a diagnostic tool to help teachers focus on learning gaps that need to be addressed with individual students.

Critics of the MAP Growth assessment

The MAP Growth assessment is not without critics or detractors. A notable flashpoint in the ongoing debate over the use of standardized assessments took place in 2013 in Seattle, WA, when a grassroots boycott was formed to prevent the use of the MAP Growth assessment at Garfield High School. In their boycott, the teachers of Garfield cited several concerns: the exam being of questionable value for high school students, lack of ability for teachers to see the questions on the test (and a belief that it was not aligned with Common Core), loss of instructional time to administer the test, use of technology resources to administer the exam, and an objection to the use of the assessment as an evaluation tool (Strauss, 2013). It could be argued that this situation is fairly representative of some of the commonly cited concerns about standardized assessment generally; particularly those concerns about the assessment being used for teacher evaluation, as well asconcerns that it can harm the disadvantaged students being assessed. Peter Hendrickson, a retired testing director and original collaborator in the development of the original NWEA assessments, argued that the focus on whether or not assessments are "good" or "bad" misses the point; instead, he encouraged those wishing

to evaluate an assessment to examine whether the assessment can and is being used for the purpose for which it was designed, and whether it helps teachers and students to measure what educators want students to learn (Shaw, 2013). It would appear that the teacher's concerns at Garfield High School came, in large part, out of the perception that the assessment was being used for things it was not designed to do. NWEA's own technical report and guidance for the MAP Growth assessment states many intended purposes of the exam; none of them include teacher evaluation (NWEA, 2019). Events like what happened in Seattle are related to the wider discourse on the use of standardized assessments and help show how their use, even in alignment for the assessments intended purposes, may be challenged.

Attribution Theory and Bias

Attribution theory and bias suggest that careful validity checks will need to be conducted on the instrument used in this study to evaluate the connection between MAP Growth scores and the student's rating of self-efficacy and academic mindset. There are two particular forms of attributional bias that could prove to provide interference in the collection of data in this study: the fundamental attribution error and self-serving attributional bias. The fundamental attribution error suggests that a researcher may overemphasize the characteristics or disposition of an individual being researched rather than accredit their responses or actions to situational pressures or constraints (Ross, 1977; Skitka et al., 2002). In the context of this dissertation's research, it will be important to determine where the students' self-efficacy or academic mindset are coming from: the MAP Growth results or some other factors or conditions independent of the MAP Growth

results. People are more likely to attribute positive outcomes or events to themselves but push off negative outcomes or events onto other people or factors. Mezulis et al. (2004) conducted a meta-analysis of 266 attributional bias studies and found that the average *d* was 0.96, indicating a strong self-serving positivity bias. Furthermore, their meta-analysis found that there was a high degree of universality to this phenomenon across different age levels, genders, and cultures (Mezulis et al., 2004). This research would suggest a distinct possibility that the survey results of this dissertation's instrument could over emphasize positive perceptions of self-efficacy and academic mindset.

Gaps In Practice and Research

For a variety of contextual and historical reasons, including those already discussed, standardized testing, or anything that looks like it, can face challenges in the current climate of education. Standardized testing has been used to unfairly punish teachers for low performance, promote values that are not always strongly favored in education, unfairly punish students with areas of extra need (low SES, EL, SPED, and minoritized students), and when the tests are high-stakes, they have encouraged gaming and cheating. The state of California has transitioned away from the NCLB era and into our current testing regime of standardized testing, in which the tests have become arguably harder (with more critical reading, writing, problem solving, and no multiple-choice questions), but the model of evaluating the testing data has not evolved with the new test. This has created two very large potential gaps in practice. First, because of the history of standardized testing and its perceived abuses, many educators are not inclined to favor or take seriously the data generated by the tests. Second, because the data generated by our

current state assessments are end-of-year assessments only, and primarily compare cohort to cohort rather than student growth throughout the year, they do not provide useful instructional data to teachers. Taken together, these two gaps in practice make it difficult to encourage teachers and students to use standardized testing data as a motivational tool for the development of self-efficacy and academic mindset; difficult, but I do not think impossibly so. Much of the self-efficacy and academic mindset research reviewed tended to focus on interventions that were more psychologically or motivationally grounded, or on academic interventions that revolved around tasks and feedback on those tasks. This study aims to address this gap by exploring whether or not standardized testing of a more formative variety, like the MAP Assessment, can provide more meaningful analysis of student growth than the currently inadequate end-of-year testing, and whether this formative testing data could help contribute to increased self-efficacy and academic mindset among students. If it could, then it may be possible to turn a historically unmotivating educational practice (standardized testing) into a practice that, instead, contributes to the development of positive self-efficacy and academic mindset, which much research has shown to be powerful drivers of student achievement.

Chapter III: Methodology and Research Design

In this chapter the purpose of this research, the research questions, rationale of the design, and design of the study will be explored. The population and sample, instrumentation, data collection design, and analysis techniques will be demonstrated. Lastly, the limitations of the study will be considered.

Statement of the Problem

Math achievement is a major determinant of student college and career readiness both in the state of California and nationally. Unfortunately, math performance suggests that the system is not adequately preparing students. Statewide in California, only 39.73 percent of students met or exceeded standard in mathematics (across all grade levels and subgroups tested) (California Department of Education, 2020).

Several districts within the state of California, collectively known as the CORE districts, have been focusing on noncognitive factors as an area of study to help understand student levels of achievement and help monitor continuous school improvement. The CORE districts are currently using the MESH survey as one of their primary instruments for assessing four noncognitive indicators self-management, self-efficacy, growth mindset, and social awareness. Their criteria for using these noncognitive factors is their belief that they show the most promise for meaningfulness, measurability, and malleability (Transforming Education, 2016). Many schools and districts have large numbers of students operating at a low achievement level, creating the potential for high growth, but these schools and districts are challenged to realize this opportunity because of poverty, systematic racism, and institutional barriers. The

growing interest in noncognitive indicators likely stems from acknowledgement that these indicators are more likely to be changed through intervention than the various social, economic, and political forces that impact the students.

There are many within the educational system that loath standardized testing due to a wide range of historical, social, political and economic arguments. This research explored potential gaps in practice that have developed because of the animosity towards standardized testing, namely: Can standardized testing be pragmatically used to help develop noncognitive factors, which the research has been shown to be an important determinant to student achievement?

Purpose of the Research

This study had three purposes: (1) determine the relationship, if any, between noncognitive factors and mathematics for at promise middle school students and determine the relationship, if any between self-efficacy (SE) and/or academic mindset (AM) and mathematics achievement for at promise middle school students; (2) study which factors predict math achievement; and (3) determine the reliability of the developed SEAM survey to measure self-efficacy and academic mindset, as well as measure the SEAM survey's validity.

Research Methodology and Study Design

The original proposed study was a primarily quantitative design that drew on strengths from both qualitative and quantitative research designs. The intended mixed methods design was chosen in alignment with my pragmatic worldview, attempting to address problems through pluralistic, real-world practice oriented and focused on the

consequences of actions (Creswell & Guetterman, 2019). However, in consultation with my committee, and considering the added complexity of interviewing students during the COVID-19 pandemic, the study was scaled back to an explanatory correlational quantitative design, using survey research and secondary data. My research questions seek to understand the relationships between three sets of variables: demographic variables (gender, ethnicity, socio-economic status, and English learner status), students' self-efficacy ratings, and students' academic mindset ratings. Creswell & Guetterman (2019) describe the advantage of correlational research when the goal of the study is to relate variables to each other and see if they influence each other. Two of my variable sets are related to students' own beliefs or attitudes. Survey research is a widely used methodology to attempt measurement of beliefs or attitudes, particularly when a large sample prospective population is targeted for study (Creswell & Guetterman, 2019; Fowler, 2009). The student's self-efficacy and academic mindset ratings were collected from the SEAM survey which I adapted from the MESH survey. The types of research questions in this study are supported by correlational studies and survey research (Creswell & Guetterman, 2019; Fowler, 2009).

Research Methods

The study began with the collection of data from the fall 2020 administration of the NWEA MAP Growth assessment. After the assessment was administered, students responded to the SEAM survey. While participating in the SEAM survey, students were in possession of their own student profile page, which gave them their recent and historical score reports. After students completed the SEAM survey, their responses were

combined with their testing and demographic data. Once the data were combined, I used

SPSS to run descriptive, correlation, regression, and a Wright analysis to address the

research questions.

Study Population

The participants in this study were 8th graders attending traditional public middle or

junior high schools serving 6th-8th graders in one target district. The target district had six

middle and junior high schools that fit these criteria, each with varying numbers of 8th

grade students. There were a total of 1355 8th grade students throughout the six school

sites. Table 7 provides a detailed description of the sample populations and schools that

were studied in this research.

Table 7

Target District's School Characteristics

Schools	Low SES	English Learners	Ethnicity
School A N= 311	6.5%	10.7%	African American – 1.6% Filipino – 1.3% Asian – 1.3% Latinx – 39.1% White – 54% American Indian – 0.6% Pacific Islander – 0.6%
School B N= 211	94.9%	34.3%	African American – 0.2% Filipino – 0.2% Asian – 0.3% Latinx – 95.2% White – 4.1% American Indian – 0.2%
School C N= 205	92.7%	47.6%	African American – 0.1% Filipino – 0.9% Asian – 0.3% Latinx – 95.9% White – 2.2% American Indian – 0.3% Pacific Islander – 0.3%
School D N= 236	94.6%	48%	African American – 0.3% Filipino – 1.4% Asian – 0.3% Latinx – 92.5% White – 5.5%
School E N= 158	96.2%	39.4%	American Indian – 0.4% Latinx – 96.6% White – 3%
School F N= 234	94.3%	50.4%	Filipino – 0.2% Asian – 0.8% Latinx – 96.4% White – 2.5% American Indian – 0.2%

Students chosen for this proposed study came from a randomly generated sample of students who provide consent for the study.

Study Sampling

To understand the relationship between noncognitive factors, my SEAM survey, and math achievement, all 8thgraders in the district were targeted for the study. The researcher sought approval from district personnel and also reached out to the administrations of each middle school or junior high targeted for the study. Each administration sent out promotional materials through their communication channels (School Messenger, parent emails, and social media). Sampling was ultimately conducted by personal email invite from the researcher to parents first, for consent (see Appendix A), and then to students, for assent (see Appendix B). Students were first invited by email, but, given a low initial response rate, follow up phone calls, letters, and messages were sent to the prospective pool of students. The researcher's random sampling generated 56 possible student participants; 45 opted to participate in the study, two moved after parent consent but before the study took place, and nine opted out. While the sample was intended to reflect the demographics of the district overall, it still experienced some skewing. Due to pandemic restrictions, all parental consent had to be secured virtually, via DocuSign and email. As a result, schools with larger numbers of parents who had submitted email addresses to their students' schools were sampled. Table 8 provides a breakdown of the possible and actual student participants.

Table 8

Possible & Actual Student Participants

Schools	Possible Participants	Actual Participants
School A N= 311	280	27
School B N= 211	70	8
School C N= 205	81	4
School D N= 236	58	4
School E N= 158	17	1
School F N= 234	67	1

As a result of the pandemic, and the resultant consent procedure restrictions, more than half of the sample came from the school with the highest percentages of parents that gave an email address to their student's school.

Description of Setting

Data collection took place during the beginning of the 2020-21 school year as the United States and world were deep in the throes of a global pandemic. During the time of the assessment administration, students in the target district were being virtually schooled from home. Virtual instruction began in the spring of the 2019-20 school year and continued into the 2020-21 school year. Data collection was done online after the

students finished the fall MAP Growth assessment. The fall MAP Growth assessment was started during the second week of school and lasted for three weeks. Students received the survey by email the week after the fall MAP Growth assessment window closed and it was sent several more times over the next several weeks. Multiple survey solicitations were sent out because the researcher found that many students were not seeing the survey link, overwhelmed by the far higher-than-average volume of digital communication necessitated by virtual learning.

Instrumentation

Two instruments were used to collect data for this study: the MAP Growth assessment and the SEAM survey. The SEAM survey was an adaptation of the MESH survey created by the researcher to address the research questions of the study.

SEAM Survey Design. The SEAM survey design was informed by the widely used MESH survey; its adaptations were guided by best practices of education measurement and assessment. The adaptations were guided by the NRC Assessment Triangle and Mark Wilson's Four Building Blocks of instrument development (Pellegrino et al., 2001; Wilson, 2005). In the NRC's Assessment Triangle, efficacious instrument design requires careful attention to three core areas: cognition, observation, and interpretation. In Wilson's approach, there are four areas to pay attention to: construct maps, items design, outcome space, and measurement models. Items one and two of the NRC Assessment Triangle and items one and two of Wilson's 4 Building Blocks overlap. With regard to

the Assessment Triangle's third corner, interpretation, it is subdivided into two categories within Wilson's Building Blocks: outcome space and measurement models.

Construct Maps. Wilson's first building block is the construct map. Construct maps are designed and created to provide a model of the construct being measured and to ensure a wide range of variance in the instrument's design (Wilson, 2005). For this study, two construct maps were created: one to measure self-efficacy and another to measure academic mindset. The construct maps for this study were developed in an iterative process involving research on the literature of academic mindsets and self-efficacy, a rudimentary pilot study, and a four-course workshop involving my advisor and other doctoral students, (see Appendices C & D).

Items Design. Wilson's second building block, item design, revolves around the format and structure of the instrument and how it explicitly connects with the construct(s) being surveyed/tested. The SEAM survey, developed by the researcher, had three key parts, each designed to address one of the research questions in the study and the two constructs.

Survey development started with a rudimentary pilot during the fall of 2018 and spring of 2019, in which I conducted a rough test of some basic self-efficacy questions in a pre-test/post-test structure after students took a MAP Growth assessment. Preliminary findings from this survey spurred further interest in simple interventions related to self-efficacy and academic mindset. Over the next year several instruments to track self-efficacy and growth mindset were investigated, before arriving at the MESH survey. Selection and refinement of the items that would eventually make up the proposed SEAM

survey came by way of construct maps developed through workshop work with my advisor and several other members of various cohorts during this spring; see Appendices B & C.

The MESH survey was developed by the CORE districts to assess noncognitive indicators. The MESH survey was chosen as the base instrument to develop this dissertation's SEAM survey due to a few key factors: it was developed in consultation with the CORE districts and leading researchers in the field of noncognitive indicators, it was widely field tested, and it had strong evidence for validity and reliability. The MESH survey in its original form has 25 items: 9 items on self-management, 4 on academic mindset, 4 on self-efficacy, and 8 on social awareness; see Appendix E (Transforming Education, 2016).

The original MESH survey was then adapted and the three-part SEAM survey was developed. The first part of the survey explores the academic mindset of the students (RQ1 and RQ2). It is comprised of 8 questions: 4 non-subject specific academic mindset questions and 4 math specific academic mindset questions. All of the questions were modified from the original construction of the MESH survey questions to reflect a positive growth mindset orientation as opposed to the negative growth mindset orientation in their original construction.

The second part of the survey explores academic mindset and self-efficacy, in connection with performance on the MAP Growth assessment (RQ1 and RQ2). These six questions are original questions, not adapted from the MESH survey, and their purpose is

to identify possible connections between students' self-efficacy or academic mindset ratings and their performance on the MAP Growth assessment.

The third section of the survey explores student self-efficacy (RQ1 and RQ2). Similar to the first section of the survey, it is comprised of 8 questions: 4 non-subject specific self-efficacy questions and 4 math specific self-efficacy questions. All of the questions were modified from the original construction of the MESH survey questions to reflect a positive growth mindset orientation, as opposed to the negative growth mindset orientation in the original construction. All major refinements of the SEAM survey can be found in Appendix F & G. All items for the final iteration of the SEAM survey can be found in Appendix H.

The SEAM survey went through an expert panel review and student think aloud protocols. While the MESH survey was vetted for validity and reliability, my adaptations have made validity and reliability protocols invalid. Reviews by an expert panel consisting of a middle school teacher, middle school principal, and county office of education administrator helped to further refine the instrument. Student think-aloud protocols, including work with two middle school students, led to further refinement of the SEAM survey items to ensure all adaptations would still be well understood by students. Lastly, it was one of my committee members who encouraged me to frame the instrument questions with positive mindset; the original MESH questions and my original adaptations were not framed this way initially.

Data Collection Procedures

All potential student respondents and their parents were pre-notified of the study by email and in the summer newsletters of each participating school (in consultation with their administration). In addition, promotional materials were distributed to each of the sites to help raise awareness, and math teachers at each site were briefed about the study. During the first week of school, the consent and assent forms were emailed and reminder emails and phone calls went out to the parents.

Participants completed the SEAM survey after completing the fall administration of the MAP Growth assessment. It was expected that the survey would take 10-20 minutes; the average time spent by respondents was 8 minutes. While completing the SEAM survey, they had their MAP Growth family profile report for reference; this is a report that shows them their score history, assessment scores, and percentiles from the exam they just completed. The survey was uploaded into Qualtrics for students to complete. The researcher realized early on that the response rate was slow as a result of students being inundated with emails during distance learning. Follow-up phone calls, texts, letters, and messaging were sent to increase the response rate. The overall response rate was 3.32% of all eligible participants. Upon student completion of the survey in Qualtrics, the data was cleaned in Excel and moved to SPSS for data analysis.

Demographic data and test scores were collected via secondary sourcing. As a district employee, and with district permission, I was able to collect recent test scores, growth scores and demographic data (EL level, SES, ethnicity, and gender) from our internal databases and merge that data with the results from the survey.

Data Analysis

This quantitative study relied on a few different data analysis strategies. Table 9

presents a list of the data analysis strategies that were used in the study.

Table 9

Data Analysis Summary

Research Question	Analysis	Model/Method	Software
RQ1, RQ1a, RQ1b Relationship between noncognitive factors, self-efficacy, academic mindset, and math achievement.	Quantitative analysis of frequency of responses and correlation of noncognitive factors.	Descriptive statistics Correlation Analysis	MS Excel SPSS 27
RQ2 Factors that predict math achievement.	Quantitative analysis of relationships between, noncognitive factors, demographics, and RIT scores.	Descriptive statistics Regression Analysis	MS Excel SPSS 27
RQ3 Reliability and Validity of SEAM's measurement of SE and AM.	Quantitative analysis of survey instrument	Factor Analysis and Wright Maps	MS Excel SPSS 27 ConQuest

When data collection was completed, descriptive statistics were run in SPSS to observe

the frequencies of the various demographic items as well as each of the survey instrument

items. Correlational analysis was run between the noncognitive items in the survey and

the student's most recent RIT scores. Next, a regression analysis was run between those

noncognitive items in the survey, the student's demographic factors, and their RIT scores.

Evidence for Validity and Reliability

Validity and reliability checks are crucial in the creation, adoption, or adaptation of an instrument. Without understanding the validity or reliability of an instrument, a researcher or policy maker could risk putting an instrument into the field that is not measuring what they think it is, then make inferences or implement policies based on that flawed data. In the context of school testing, high stakes tests can be differentiated from low-stakes tests as those that are used for personnel decisions, inform accountability measures, or generate awards. I would propose that any instrument, whether it be a test or not, that may be used to influence budgets, professional reputations, and perceptions of a school or program has the potential for high stakes capacity. It is for this reason that one of the leading guidelines on validity and reliability suggests the importance of working diligently to minimize measurement errors and studying carefully how the instrument was constructed to ensure its validity and reliability (AERA et al., 2014). This study employed some measures to collect evidence of validity and reliability and to reduce, where possible, any threats to that validity and reliability.

Evidence for Validity. Validity evidence is how a researcher or policy make can make claims about what their instrument is measuring or doing. According to AERA et al. (2014), "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses" (p. 11). This makes the collection of validity evidence critical in the effort to ensure that tests or instruments are doing what they say they are doing and that they are being used for their designed and intended purposes. The Standards for Education and Psychological Testing recognizes five types

of validity evidence: content validity, response process validity, internal structure validity, relations to external variables validity, and consequential validity (AERA et al., 2014). This study was able to collect validity evidence for content, response processes and internal structure but unable to collect validity evidence for relations to external variables or consequences.

Content Validity. Content validity evidence allows a researcher to examine the relationship between the content or purpose of the instrument and the construct being measured; this process is critical in the development of an instrument. In education, this process of "alignment" between student learning standards and test content applies to the process engaged in during this study, aligning the constructs (self-efficacy and academic mindset) to the instrument (AERA et al., 2014). The SEAM instrument was developed with several content validity strategies in mind. Construct maps were developed for academic mindset and self-efficacy and allowed the researcher to provide a framework for the constructs measured (see Appendices B & C). These construct maps were informed by the literature on self-efficacy and academic mindset as well as a rough pilot study conducted the previous year. The SEAM survey went through several iterations after being reviewed by an expert panel consisting of a middle school teacher, middle school administrator, and county office of education alternative education administrator. This review helped improve the alignment between the item design and construct maps.

Response processes validity. Response process validity evidence gives a researcher an opportunity to explore the cognitive processes assumed to be taking place with the use of an instrument; this gives the researcher a chance to determine, from the user of the

instrument, what they interpreted in the instrument (AERA et al., 2014). In this study, the researcher was able to gather some response process validity evidence during two separate sessions in which the researcher asked middle school students to take the survey and report out what they were thinking during the use of the instrument. These sessions led to some slight changes to the language of the instrument to ensure that the intended constructs were probed.

Internal structure validity. Internal structure validity explores the connections between the construct and instrument item. This is important validity evidence to collect as it allows a researcher to know the degree to which the instrument is measuring what it is intended to measure (AERA et al., 2014). This study employed a Rasch item response model and Wright maps in order to verify that the construct map aligned with the survey's items design (see Appendix I for the Wright Maps created for this study).

Relations to external variables and consequences. Neither evidence for external variables nor consequences were collected for this study. External validity evidence attempts to connect other instruments that measure similar constructs to the construct of the primary instrument in order to see if they get comparable responses. If the primary instrument is in fact measuring what it thinks it is measuring, then it stands to reason that another instrument measuring the same construct should pick up the similar results (AERA et al., 2014). The purpose of consequence validity evidence is to explore the interpretations and evidence derived from an instrument in order to determine whether or not the instrument is aligned with its intended use (AERA et al., 2014). It would certainly be beneficial in follow up research to compare this instrument and its modifications to

other self-efficacy or academic mindset instruments to collect external validity evidence. Consequence validity evidence was not collected due to a lack of time to conduct such protocols; however, before this instrument could be more widely, used these protocols should be completed.

Evidence for reliability. Reliability evidence refers to "the consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported" (AERA et al., 2014, p. 33). Or to put it another way: Does the researcher get the same results with the instrument, regardless of the measurement index or tool used? If an instrument is not reliable, it would certainly impact a researcher's ability to use the instrument in future research and would call into question the soundness of the data generated. There are four types of reliability evidence that can be collected: internal consistency, alternative forms, test retest and rater reliability. Another major factor that can reduce the reliability of an instrument is random errors; in this section I will also review the random errors that were possible in this study and what was done to mitigate errors.

Internal consistency. Internal consistency data was collected to explore and understand the reliability of the instrument. Internal-consistency coefficients are one of the three recognized broadly categorized methods of exploring reliability and includes "the relationships/interactions among scores derived from individual items or subsets of the items within a test, all data accruing from a single administration" (AERA et al., 2014, p. 37). This study generated Cronbach's alpha and Person separation coefficients to document the reliability of the instrument. It is recognized that having multiple reliability

coefficients can be useful, as they have different potential sources of error and can convey different angles of information (AERA et al., 2014).

Alternate forms, test retest and rater reliability. There are three types of reliability evidence that were not collected in this study: alternate forms, test-retest, and rater reliability. Reliability evidence regarding alternate forms should be collected if there is more than one version of the survey with interchangeable parts. This allows a researcher to compare the means, standard deviations, or reliability co-efficient of those versions to show the reliability of the instrument (AERA et al., 2014). Only one version of the survey was generated, so this type of reliability evidence was not possible to collect. Test-retest involves having the person completing the instrument take it more than once, followed by study of the variations in their responses (AERA et al., 2014). This evidence was not collected because the researcher did not deem it realistic to persuade 8th grade adolescents to take the survey more than once; however, it would be certainly interesting to see how the survey results may have been different if the same students were to take the survey again while not engaged in distance learning. Finally, rater reliability entails studying the differences in scoring during qualitative analysis of an instrument when different raters are involved (AERA et al., 2014). As the instrument was a quantitative survey and was automatically scored, this evidence was not possible to collect.

Random Errors. Random errors can impact the reliability of an instrument when internal or external factors result in unpredictability regarding how respondents use the instrument in question (AERA et al., 2014). There were many conditions that could have

created random errors in this study, including the pandemic, virtual classroom fatigue, motivation, access, and distractions.

To address these random errors, respondents were pre-notified of the study via advertisement, email, and phone call. The survey was designed to be relatively short and easy to complete, using kid-friendly language, so that students would be less likely to quit due to boredom or confusion. Students were also appealed to individually with personal emails, letters, and follow ups to convey to them the importance of their responses.

Another strategy useful in minimizing impact to instrument data from random error is getting the largest sample size possible. Unfortunately, the sample size generated by this study was not ideal, but much effort was put into getting the largest possible sample from the relatively small pool available under the pandemic-imposed conditions. This included follow up emails, letters, phone calls to students and parents, and other communications to potential participants.

Ethical Considerations

Research conducted on any individual comes with important ethical considerations; research conducted on minor students comes with even more responsibility. This study was submitted to the San José State Institutional Review Board in the spring of 2020 and was granted approval during the summer of 2020 (Protocol #20165, see Appendix J). In order for any student to participate, they and their parents needed to provide assent and consent, respectively. Parental consent was obtained directly through emails to the parents and signed with DocuSign. Student assent was collected by the first question in the survey which provided students with their rights, explained the purpose of the study,

and asked for their assent. Great effort was put into ensuring the privacy of the

participants during the study, to ensure that identities were not revealed. This involved

deidentifying survey responses to ensure that the researcher was the only one who could

identify the students and only reporting data in the aggregate. The MAP Assessment is

run three times a year for the students in my district. Students are very familiar with the

assessment and the study design had relatively minimal impact on the instructional time

of the teachers and students that participated. However, one area of ethical concern was

the issue of how to ask questions about student's self-efficacy and academic mindset, in

the context of standardized testing, without unleashing the theorized forces of stereotype

threat (Steele, 1997; Walton & Cohen, 2007; Walton & Spencer, 2009). While issues of

race, socio-economic status, and language level will be studied in their relation to self-

efficacy and academic "growth" mindset, it is certainly not the goal of this study to

activate in any student feelings of inadequacy or lack of ability, especially as a result of

those factors. Through the expert and student panel reviews, piloting, and consultation

with my advisor, I believe I was able to mitigate this. One key mitigation technique was

to not bring up any of those demographic factors in the survey. Instead, those were

collected separately from the study and merged with the survey responses before the data

were deidentified.

Limitations and Significance

The biggest potential limitation that may have impacted research design and

execution was the COVID-19 global pandemic. As a result of the pandemic, schools were

shut down in spring 2020 and the spring administration of the MAP Growth assessment

was cancelled. The MAP Growth assessment is most accurate when it is run on regular intervals, and we really didn't know what the fall of 2020 would look like. This has been considered during the research design process.

Some researchers have argued that attributional ambiguity, or the idea that positive or negative feedback might not be trusted by stigmatized groups, presents the possibility that the test scores, good or bad, may not have an effect on the student's feelings of self-efficacy or academic "growth" mindset (Mendoza-Denton et al., 2008; Mendoza-Denton et al., 2010).

The potential significance of the study is that a new proposed instrument will be developed that can assess student feelings of self-efficacy and academic "growth" mindset in relation to math achievement. The study also has potential to show the viability of alternative academic measures, like the MAP Growth assessment, to be more useful for vulnerable students and more useful for their teachers.

Chapter IV: Findings

The purpose of my study is to explore, at the middle school level, the impact of academic achievement on a student's academic mindset and self-efficacy as well as the impact of a student's academic mindset and self-efficacy on their academic growth. Furthermore, the purpose of this study is to evaluate the impact of presenting students with their own academic growth data from norm-referenced interim standardized assessments on academic mindset and self-efficacy. This type of feedback could be particularly valuable for students whose personal or schooling context masks the reality that their growth is more pronounced than they perceive. To this end, three questions were addressed. First, what is the relationship, if any, between noncognitive factors and mathematics achievement for at promise middle school students, particularly regarding academic mindset and self-efficacy? Second, what factors predict math achievement? Third, how reliably and validly does the SEAM instrument measures academic self-efficacy and academic mindset?

Summary of Results Obtained from Sample

The study sample was made up of 8thgraders who had parent permission to participate in the study and also provided their assent. A total of 56 parents and guardians provided permission for the study, 45 students chose to participate, 2 moved before they had a chance to participate, and 9 opted not to participate in the study. Participants came from 6 middle schools in a single district. Table 10 presents a summary of the demographic data of the student participants.

Table 10

Survey Participant Demographics

Participant Demographics	Frequency	Percent
Gender		
Female	23	51.1%
Male	22	48.9%
Ethnicity		
Latinx	17	37.8%
White	23	51.1%
Chinese	2	4.4%
Asian Indian	3	6.7%
English Learner Level Simple		
English Only	32	71.1%
All Other EL Status	13	28.9%
English Learner Level Expanded		
English Only	32	71.1%
English Learner	3	6.7%
Redesignated Fluent English Proficient	2	4.4%
Initially Fluent English Proficient	8	17.8%
Socio Economic Status		
Free and Reduced Lunch	23	51.1%
Not Free and Reduced Lunch	22	48.9%

As Table 10 shows, the sample of students was almost evenly split between male and

female, 51.1% and 48.9% respectively. The students were also almost evenly split

between white and all other groups: white students were 51.1%, Latinx students were

37.8%, and the last two groups, Chinese and Asian Indian, were 4.4% and 6.7%,

respectively. The majority of students in the sample, 71.1%, were "English only" in their

English learner level. Students were also almost evenly split in their socio-economic

status: 51.1% of participants met criteria for free and reduced lunch while 48.9% did not.

As mentioned previously, the study population did not represent all 6 schools in the district proportionally. Because of the pandemic, and the requirement to solicit consent and assent only by virtual methods, the schools that had the most parents provide emails were sampled higher; see Table 5. One school contributed 27 out of 45 student participants, while the other 5 schools combined contributed the other 18 participants.

Results Obtained from the SEAM Instrument Responses

The SEAM instrument contained 11 questions on academic mindset and another 11 questions on self-efficacy. The SEAM instrument had two parts, each part (self-efficacy and academic mindset), contained four questions taken from the original MESH survey (with modifications), four math specific questions adapted from the original language of the MESH survey, and three questions pertaining specifically to students' MAP Growth results.

Survey Results from Academic Mindset Domain

In the academic mindset domain, there were four questions related to general academic mindset (AM 1-4), four questions related to academic mindset as it pertains to math (AM 5-8), and three questions connecting students' performance on the MAP Growth assessment with academic mindset (AM 9-11). Table 11 contains a breakdown of the responses for the academic mindset domain: including item response percentages, means, and standard deviations.

Table 11

Academic Mindset Item Response Percentages, Means and Standard Deviations

Item	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Missing Values	Mean	SD
AM1	0	0	3	26	14	2	2	1.05
	0%	0%	6.7%	57.8%	31.1%	4.4%		
AM2	0	2	7	23	12	1	2.07	.99
	0%	4.4%	15.6%	51.1%	26.7%	2.2%		
AM3	0	0	5	16	22	1	1.70	.95
	0%	0%	11.4%	36.4%	50%	2.3%		
AM4	0	3	13	21	7	1	2.36	.98
	0%	6.7%	28.9%	46.7%	15.6%	2.2%		
AM5	1	1	2	22	19	0	1.73	.84
	2.2%	2.2%	4.4%	48.9%	42.2%	0%		
AM6	0	3	10	19	12	1	2.18	1.05
	0%	6.7%	22.2%	42.2%	26.7%	2.2%		
AM7	0	1	6	14	23	1	1.76	1.03
	0%	2.2%	13.3%	31.1%	51.1%	2.2%		
AM8	1	17	20	5	2	1	4.18	.98
	2.2%	37.8%	44.4%	11.4%	4.4%	2.2%		
AM9	0	4	11	19	9	2	2.40	1.18
	0%	8.9%	24.4%	42.2%	20%	4.4%		
AM10	0	1	12	25	4	3	2.49	1.14
	0%	2.2%	26.7%	55.6%	8.9%	6.7%		
AM11	0	0	10	20	11	4	2.33	1.35
	0%	0%	22.2%	44.4%	24.4%	8.9%		

As Table 11 shows, there was a high degree of reported agreement on most of the academic mindset items; however, the table also demonstrates a non-normative distribution with 10 of the 11 items having 60% or more of the students agreeing or strongly agreeing to related items. Over 80% agreed or strongly agreed that their smartness was something that they could change (AM1), that there are many new academic things they can learn (AM3), that their math learning can improve with effort (AM5), and that there are many things in math they are capable of learning (AM 7). Between 70% and 80% of the students agreed or strongly agreed that challenging

themselves on difficult work would help them more (AM2), that they could do well on academic subjects regardless of their natural starting ability (AM4), and that their NWEA "growth overtime chart" helps them see their growth taking place (AM11). Between 60% and 70% of the students agreed or strongly agreed that working on challenging problems in math would help them learn more (AM6), that their MAP Growth percentile makes them feel like they are growing in math ability (AM9), and that their RIT score helps them understand their growing math ability (AM10). One item, "If I am not naturally smart in math, I will never do well at it," (AM8) had the largest percentage of neutral responses at 40%. This item was the math specific item that corresponded to AM4, and the difference in responses suggests that either AM8 or AM4, or both, might have confused the students. Overall, a majority of the students, on a majority of the items, self-reported by agreeing or strongly agreeing with statements that suggest they have feelings of academic mindset.

Survey Results from Self-Efficacy Domain

In the self-efficacy domain, there were four questions related to general self-efficacy (SE4-7), another four questions relating to self-efficacy as it pertains to math (SE8-11), and three questions connecting students' performance on the NWEA MAP assessment and self-efficacy (SE1-3). Table 12 contains a breakdown of the responses for the self-efficacy domain, including item response percentages, means and standard deviations.

Table 12

Self-Efficacy Item Response Percentages, Means and Standard Deviations

Item	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Missing Values	Mean	SD
SE1	0	2	9	25	8	1	2.2	.94
	0%	4.4%	20%	55.6%	17.8%	2.2%		
SE2	0	2	12	24	6	1	2.31	.93
	0%	4.4%	26.7%	53.3%	13.3%	2.2%		
SE3	0	1	12	22	7	3	2.42	1.20
	0%	2.2%	26.7%	48.9%	15.6%	6.7%		
SE4	0	2	5	21	16	1	1.93	1.01
	0%	4.4%	11.1%	46.7%	35.6%	2.2%		
SE5	0	1	12	22	9	1	2.20	.94
	0%	2.2%	26.7%	48.9%	20%	2.2%		
SE6	0	2	9	22	11	1	2.13	.99
	0%	4.4%	20%	48.9%	24.4%	2.2%		
SE7	0	2	9	24	8	2	2.29	1.1
	0%	4.4%	20%	53.3%	17.8%	4.4%		
SE8	0	1	12	18	14	0	2.00	.83
	0%	2.2%	26.7%	40%	31.1%	0%		
SE9	0	1	14	23	7	0	2.20	.73
	0%	2.2%	31.1%	51.1%	15.6%	0%		
SE10	1	6	13	19	5	1	2.6	1.07
	2.2%	13.3%	28.9%	42.2%	11.1%	2.2%		
SE11	0	1	8	25	9	2	2.2	1.08
	0%	2.2%	17.8%	55.6%	20%	4.4%		

As Table 12 shows, there was less reported agreement on several of the self-efficacy items compared to the academic mindset items. Similarly, though, the self-efficacy items also demonstrate a non-normative distribution, with 10 of the 11 self-efficacy items having over 60% of the students either agreeing or disagreeing to the individual items. Over 80% agreed or strongly agreed that their smartness was something that they could change (SE4). Between 70% and 80% of the students agreed or strongly agreed that their MAP Growth percentile demonstrated that they could succeed at math (SE1), that they can gain an understanding of difficult topics in all of their classes (SE6), that they can

meet the learning expectations of their teachers (SE7), that they can earn an A in their

math class (SE8), and that they can meet the expectations of their math teacher (SE 11).

Between 60% and 70% of the students agreed or strongly agreed that their RIT scores

shows that they succeed at math (SE2), that their MAP Growth "growth over time chart"

shows that they can succeed at math (SE3), that they can do well on tests even when they

are difficult (SE5) and that they can do well on math tests even when they are difficult

(SE9). Overall, a majority of the students, on a majority of the items, self-reported by

agreeing or strongly agreeing with statements that would suggest they have feelings of

self-efficacy.

Descriptive Statistics for MAP Growth Assessment

Math achievement in this study was measured through the student's RIT score on the

fall administration of the MAP Growth assessment. Table 13 contains the descriptive

statistics for the sample's RIT scores.

Table 13

RIT Score Descriptive Statistics

	N	Min	Max	Mean	Std. Deviation
RIT	45	196	270	231.58	17.669

The range of RIT scores in the sample was 196 to 170. The mean RIT score in the sample

was $M = 231.58$, $SD = 17.669$. The RIT scores can be further broken down by each

domain tested in the MAP NWEA assessment. The MAP Growth assessment had four

domains tested: operations and algebraic thinking, statistics and probability, real and

complex numbers systems, and geometry. Table 14 contains the descriptive statistics for these domains of the sample students.

Table 14

RIT Scores Descriptive Statistics by Domain

	N	Min	Max	Mean	Std. Deviation
Operations & Algebraic Thinking	45	192	270	232.76	18.84
Statistics and Probability	45	193	269	230.16	19.13
Real and Complex Numbers Systems	45	191	263	234.98	15.97
Geometry	45	188	281	227.89	18.02

As Table 14 illustrates, each of the four domains tested relatively similarly with comparable means across all four domains. However, the greatest standard deviation was in statistics and probability, suggesting it has the greatest variability in tested level.

Results for Research Question One

The first research question—what is the relationship, if any, between noncognitive factors and mathematics achievement for at promise middle school students—explored the correlation between noncognitive factors and math achievement. More specifically, I wanted to explore the relationship between academic mindset and math achievement (as measured by the RIT score), and self-efficacy and math achievement (as measured by the SEAM instrument). This question was explored through correlation analysis.

Correlation of Academic Mindset and Math Achievement

Correlating the relationship between academic mindset and math achievement (as measured by RIT) had mixed results. Table 15 contains a report of the correlation analysis run between academic mindset (AM) and math achievement (RIT).

Table 15

Correlation of Academic Mindset (AM) and Math Achievement (RIT)

		M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1	AM1	1.93	1.05												
2	AM2	2.07	0.99	.64**											
3	AM3	1.7	0.95	.69**	.71**										
4	AM4	2.36	0.98	.05	.30*	.01									
5	AM5	1.73	0.84	.68**	.46**	.53**	.15								
6	AM6	2.18	1.05	.40**	.54**	.40**	.36*	.31*							
7	AM7	1.76	1.03	.43**	.22	.51**	-.09	.53**	.29*						
8	AM8	4.18	0.98	-.1	-.04	-.14	-.28	-.41**	-.27	-.2					
9	AM9	2.4	1.18	.24	.17	.33*	.27	.23	.25	.54**	-.14				
10	AM10	2.49	1.14	.07	.01	.17	.04	.09	-.04	.49**	.	.38*			
11	AM11	2.33	1.35	.11	.09	.2	.06	.1	.2	.11	-.05	.2	.41**		
12	RIT	231.6	17.7	.30*	.21	.30*	-.07	.32*	.03	.07	-.16	.2	-.09	.34*	

*N = 45. * p < .05; ** p <.01*

Table 15 demonstrates that the SEAM instrument had a positive correlation at the $p <$ *.05* with academic mindset instrument items 1, 3, 5, and 11. Academic mindset instrument item #1, AM1 $r = .30$, $p = .05$, "my smartness is something that I can change, if I try at school," showed strong positive correlation to math achievement. Academic mindset instrument item #3, AM3 $r = .30$, $p = .05$, "there are many new academic things I can learn," showed strong positive correlation to math achievement. Academic mindset instrument item #5, AM5 $r = .32$, $p = .04$, "my math learning is something that I can improve with effort," showed strong positive correlation to math achievement. Lastly, academic mindset instrument item #11, AM11 $r = .34$, $p = .02$, "the NWEA 'growth over time chart' helps me understand how I am improving in my understanding of math," showed strong positive correlation to math achievement. Other positive correlations between math achievement and academic mindset at statistically insignificant levels were

detected in items AM2, AM6, AM7, and AM9. These refer to items: "challenging myself on difficult school work will help me learn more," "working on challenging problems in math will help me learn more," "there are many things in math I am capable of learning," and "my NWEA growth percentile makes me feel like I am growing in my math ability." Negative correlations between math achievement and academic mindset were detected in items AM4, AM8 and AM10. These refer to items: "I can do well in academic subjects I am naturally smart in and in those that might be difficult at the start," "if I am not naturally smart in math, I will never do well at it," and "my NWEA RIT score helps me understand how I am growing in my math ability."

Correlation of Self-Efficacy and Math Achievement

Correlating the relationship between self-efficacy and math achievement (as measured by RIT) had mixed results. Table 16 contains a report of the correlation analysis run between self-efficacy (SE) and math achievement (as measured by RIT).

Table 16

Correlation of Self-Efficacy (SE) and Math Achievement (RIT)

		1	2	3	4	5	6	7	8	9	10	11	12
1	SE1												
2	SE2	.34*											
3	SE3	.45**	.66**										
4	SE4	.09	.12	.14									
5	SE5	.13	.03	-.04	.61**								
6	SE6	.26	.13	.18	.67**	.77**							
7	SE7	.21	.09	.06	.63**	.67**	.69**						
8	SE8	.26	.21	.05	.27	.32*	.22	.38*					
9	SE9	.37*	.18	.27	.42**	.57**	.59**	.55**	.64**				
10	SE10	.48**	.24	.13	.29	.51**	.58**	.52**	.56**	.72**			
11	SE11	.27	.16	.14	.68**	.52**	.65**	.85**	.46**	.56**	.50**		
12	RIT	.19	.16	.29	.04	.05	.1	.08	.22	.31*	.11	.16	

N = 45. * p < .05; ** p < .01

As Table 16 demonstrates, the SEAM instrument detected statically insignificant levels of correlation between 10 of the 11 instrument items with no negative correlations. Only one item, self-efficacy instrument item #10, SE9 r = .31, p = .04, "I can do well on my math tests, even when they are difficult," showed strong positive correlation to math achievement.

Correlation of SEAM Instrument and Math Achievement

A correlation was also run between math achievement as measured by RIT and the collective ratings of academic mindset and self-efficacy measured by the SEAM instrument. A Pearson product-moment correlation coefficient was computed to assess the relationship between math achievement (as measured by RIT) and the combined ratings of academic mindset and self-efficacy (as measured by SEAM). There was a positive correlation between the two variables, r = 0.111 n = 45, p = 0.469. A scatterplot

summarizes the results (see Figure 2). Overall, there was a weak positive correlation between math achievement and SEAM. Increases in math achievement were weakly correlated with increases in SEAM ratings.

Figure 2

Scatterplot of Math Achievement (RIT) and Self-Efficacy and Academic Mindset (SEAM) Correlation
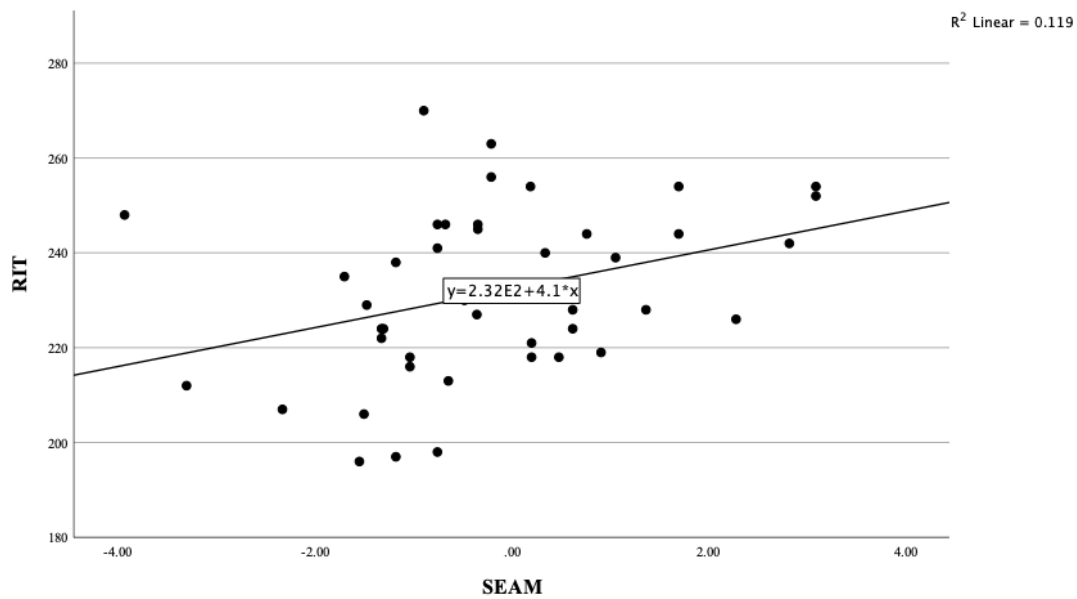
These results aligned with the correlation numbers generated from running the self-efficacy and academic mindset items individually; while there were some positive correlations, few were at a statistically significant level overall.

Summary of Research Question 1

In regard to research question #1, the data suggest that while there are some statistically significant correlations detected between some academic mindset/self-efficacy items and math achievement, it would be difficult to claim strong overall correlations. With the presence of some negative correlations and statistically

insignificant positive correlations, it would not be appropriate to argue that the SEAM instrument detected a strong correlation between math achievement, self-efficacy and academic mindset. However, the relatively small sample size of 45 also increases the difficulty of securing stable quantitative analysis from the sample.

Results for Research Question 2

The second research question, what factors predict math achievement, used a regression analysis to study the relationship between noncognitive factors, math achievement, English learner level, and socio-economic status. Based on the literature review, it was hypothesized that socio-economic status would play a major role on math achievement as measured by RIT scores. It was also expected that English learners may have a higher likelihood of struggling with math achievement as measured by RIT scores. This study also sought to determine whether or not noncognitive factors (academic mindset and self-efficacy) had a direct impact on math achievement as measured by RIT scores.

In the process of collecting validity and reliability evidence for research question three, a Rasch Partial Credit Model was generated to ensure item fit in the SEAM instrument. A unidimensional random coefficient multinomial logit (URCML) model was generated. This unidimensional Rasch Partial Credit Model generated an estimate of the amount of SEAM belief each student had on a logit scale (this is the mathematical unit used in Rasch measurement). The Expected A Posterior (EAP) of SEAM belief was one of three explanatory variables (the other two were English learner level and socio-

economic status) used in the linear regression; the dependent variable was math

achievement as measured by (RIT). Table 17 presents the regression results.

Table 17

Regression Summary for SEAM Scale

				CI 95%	
Variable	Coef.	t	SE	LL	UL
(Constant RIT)		18.127	11.619	187.149	234.077
SES*	0.548	4.274	4.482	10.104	28.207
English Learner	-0.135	-1.052	4.959	-15.234	4.796
SEAM*	0.34	3.107	1.301	1.415	6.672

N=45, Coef.=regression coefficient, SE=standard error, CI=confidence interval,
LL=lower limit, UL=upper limit, *= significant at the $p<.05$ level.

As Table 17 shows, there was a statistically significant relationship between math

achievement (RIT) and socio-economic status at the .05 level (t=4.274, CI [10.104-

28.207]). This suggests that socio-economic status is a statistically significant

determinant of math achievement as measured by RIT. There was also a statistically

significant relationship between math achievement (RIT) and the SEAM belief scale at

the .05 level (t=3.107, CI [1.415-6.672]). This suggests that noncognitive factors (self-

efficacy and academic mindset) were a moderate determinant of math achievement.

Summary of Research Question 2

Evidence from the regression analysis suggests a relationship between students'

socio-economic status and math achievement and also suggests a relationship between

students self-reported beliefs of self-efficacy and academic mindset as measured by the

SEAM instrument.

Results for Research Question 3

The third research question, how reliably does the SEAM indicator measure levels of self-efficacy and academic mindset and what is the evidence for the validity of SEAM scores to measure students, explored the validity and reliability of the SEAM instrument as modified from the original MESH survey.

Validity Evidence

The Standards for Education and Psychological Testing recognizes five types of validity evidence: content validity, response process validity, internal structure validity, relations to external variables validity, and consequential validity (AERA et al., 2014). During the course of the study, three primary forms of validity evidence were collected to validate the SEAM instrument: content validity evidence, response processes validity evidence, and internal structure validity evidence. This section presents the validity evidence findings in support of this instrument.

Content Validity. Content validity evidence allows for a researcher to examine the relationship between content or purpose of the instrument and the construct being measured. This process is critical in the development of an instrument and all other validity and reliability evidence depends on the content validity evidence (Wilson, 2005). In education, this process of "alignment" between student learning standards and test content related to the process engaged in during this study of aligning the constructs (self-efficacy and academic mindset) to the instrument (AERA et al., 2014). The SEAM instrument was developed as an adaptation of the MESH survey, a widely used measure of noncognitive factors in the state of California that is grounded in the literature. This

instrument was shown through the literature to be backed and developed by some of the leading experts in academic mindset and self-efficacy (Transforming Education, 2016). The adaptations of the MESH survey created by the researcher were reviewed by two separate expert panels, including educators at the K-12 and university level. This iterative review process of the original survey and adaptations led to the development of two construct maps, a representation of the instrument evidence: academic mindset and self-efficacy, that provided grounding for the constructs to be studied (see Appendices B & C).

Response Processes Validity. Response process validity evidence gives a researcher an opportunity to explore the cognitive processes assumed to be taking place with the use of an instrument; this gives the researcher a chance to determine from the user of the instrument what they interpreted when using the instrument (AERA et al., 2014). This process of validation allows the researcher to ensure alignment between the intended assessed constructs and how the students would interpret the instrument items. The SEAM instrument was developed in consultation with two separate middle school students, one a 6thgrader and the other an 8th grader. These two students took and responded to the survey, then I conducted a think aloud protocol with each student in which valuable response process feedback was collected. This, in combination with expert panel review, led to three different versions of the SEAM survey's development. Revisions made to the original MESH survey included creating a single outcome space continuum; strongly, agree, neutral, disagree, and strongly disagree, as students reported that having two separate outcome spaces for each part of the survey was confusing.

Another example of a type of revision made as a result of the think aloud protocols included making some items more specific. Academic mindset instrument item 2 (AM2) provides an example of this modification: "challenging myself won't make me any smarter," became "challenging myself on hard academic things won't make me any smarter." Expert panel review also pointed out the negative framing of the original MESH items and recommended positive framing. As a result, AM 2 was subsequently edited to "challenging myself on difficult school work will help me learn more."

Internal Structure Validity. Internal structure validity explores the connections between the construct and instrument items. This is important validity evidence to collect as it allows the researcher to know the degree to which the instrument is measuring what it is intended to measure (AERA et al., 2014). This study employed three forms of internal structure validity evidence: a Rasch IRT analysis, a Wright map, and a partial credit model. A Rasch IRT analysis of the SEAM scale was generated to provide evidence of whether the relationships between items and respondents was aligned with the construct. This Rasch IRT analysis yielded an empirical set of parameters in the form of a Wright map, which was generated to help represent the locations of item thresholds and locations of respondents on a single scale. Wright Maps are useful for providing evidence that the construct being measured has a range of responses and to help detect poor item fit (Wilson, 2005). A partial credit model was generated to study the

relationship of model fit for the SEAM instrument in terms of expected versus actual locations.

      The Wright Map. The Wright map was generated with ConQuest software as a representation of the unidimensional scale of SEAM. In the Wright Map, the left side represents the number of students who responded at various levels of SEAM belief on the logit scale and will be referred to as person proficiencies. The right side of the Wright map represents the probability of individual SEAM instrument items to be endorsed by students at each respective SEAM logit level and will be referred to as item thresholds. Next to each instrument item number is a .1 or .2 threshold; the .1 threshold refers to the probability of a student at a particular logit level moving from strongly disagree, disagree, or neutral to agree, and the .2 threshold refers to students moving from the agree to strongly agree level. The distribution of respondent and item locations shown in Figure 3 is on a single logit scale.

Figure 3

Wright Map of Person Proficiencies and Item Thresholds for the SEAM Scale

```
        -------------------------------------
                            |8.2
     5                      |
                            |
                            |
                            |
     4                      |
                            |
                            |
                            |
                         XXX|
     3                      |8.1
                          XX|
                            |
                          XX|
                            |21.2
     2                      |
                         XXX|4.2 10.2 20.2
                            |9.2 13.2 14.2
                          XX|12.2 16.2 18.2
                            |6.2 17.2 22.2
     1                    XX|2.2 11.2 19.2
                         XXX|
                         XXX|15.2
                          XX|1.2
                     XXXXXXX|
     0                    XX|3.2 5.2 7.2
                            |
                 XXXXXXXXXX |
                          XX|21.1
                    XXXXXXXX|
                          XX|10.1
    -1                   XXX|4.1 9.1 13.1 20.1
                    XXXXXXXX|14.1 18.1
                         XXX|6.1 12.1 16.1 17.1 22.1
                         XXX|19.1
                            |2.1 11.1
    -2                      |
                            |15.1
                          XX|1.1
                            |
                            |3.1 5.1 7.1
    -3                      |
                          XX|
                            |
                          XX|
    -4                      |
                            |
                            |
                            |
    -5                      |
                            |
                            |
                            |
        =====================================
```
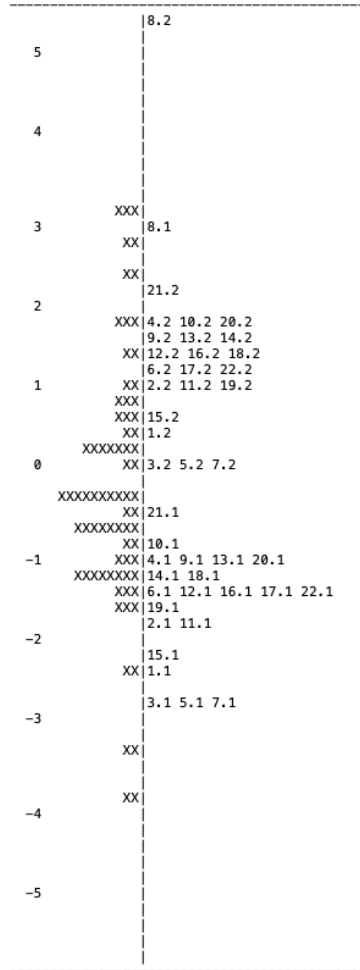
Each X on the Wright map represent 0.6 cases. For the SEAM construct of self-efficacy and academic mindset, the mean belief on the logit scale was -.2050 with a standard deviation of 1.4868 and had a range of -3.9286 to 3.0896. A somewhat bell-shaped distribution of person proficiencies suggests a fair amount of variability of SEAM levels as detected by the SEAM instrument.

100

The right side of Figure 3 shows the Thurstonian thresholds for the construct scoring categories, which in this study is defined as the SEAM belief level at which the probability of achieving that rating or higher reaches 50%. For example, when looking at the map item 21.2, ("x.y") is used to indicate the y-th threshold ("2") of the x-th item ("21"). Each step threshold is plotted at the point where the respondent has a 50% chance of achieving at least the indicated level of performance on that individual item. The range of item locations was approximately -2.8 to 3 logits.

The relationship between the left and right side of the Wright map indicates that student responses are spread across the range of item thresholds measured by the SEAM instrument. From the raw data, it can be seen that 27 of the students were negatively scaled on the SEAM scale and 18 were positively scaled. All of the 22 items are clustered in such manner that those in the agree to strongly agree threshold (.2) are positively rated on the SEAM scale, while those in the strongly disagree, disagree and neutral to agree threshold (.1) are negatively rated on the SEAM scale. One item, AM 8, "If I am not naturally smart in math, I will never do well at it," appears to be a clear outlier; both thresholds were far above almost all respondents suggesting no range or variance in answers to that instrument item.

Model Fit. Further internal structure evidence was collected through generation of a partial credit model to generate an item fit analysis. Table 18 provides the parameter estimates standard errors and mean square weighted fit and agreeability statistics for the partial credit model applied to the SEAM instrument data.

Table 18

Item Parameter Estimates, Standard Errors and Mean Square Weighted Fit and t
Statistics for the SEAM Scale

Variables				Weighted Fit	
Items	Estimate	Error	Infit Mean Square	Confidence Interval	Infit t
AM1	-0.959	0.277	0.71	(0.61, 1.39)	-1.5
AM2	-0.375	0.266	0.94	(0.62, 1.38)	-0.3
AM3	-1.358	0.284	0.99	(0.60, 1.40)	0
AM4	0.407	0.271	1.56	(0.63, 1.37)	2.6
AM5	-1.329	0.28	0.88	(0.61, 1.39)	-0.6
AM6	-0.143	0.267	1.07	(0.62, 1.38)	0.4
AM7	-1.352	0.285	1.32	(0.60, 1.40)	1.5
AM8*	*4.258*	*0.575*	*2.85*	*(0.13, 1.87)*	*3*
AM9	0.296	0.271	1.01	(0.63, 1.37)	0.1
AM10	0.42	0.278	0.82	(0.62, 1.38)	-1
AM11	-0.372	0.276	1.45	(0.61, 1.39)	2
SE1	0.025	0.267	1.05	(0.62, 1.38)	0.3
SE2	0.308	0.269	0.71	(0.63, 1.37)	-1.7
SE3	0.216	0.274	0.92	(0.62, 1.38)	-0.4
SE4	-0.813	0.272	1.17	(0.61, 1.39)	0.9
SE5	0.006	0.267	0.9	(0.62, 1.38)	-0.5
SE6	-0.142	0.267	0.77	(0.62, 1.38)	-1.3
SE7	0.072	0.269	0.8	(0.62, 1.38)	-1.1
SE8	-0.294	0.264	1.15	(0.63, 1.37)	0.8
SE9	0.36	0.267	0.62	(0.63, 1.37)	-2.3
SE10	0.887	0.279	0.89	(0.62, 1.38)	-0.6
SE11	-0.117	0.271	0.78	(0.62, 1.38)	-1.2

*Misfit item in italics.

In analyzing Table 18, the goal is to examine the residuals, the difference between the

observed and expected scores, for individual items and respondents. Through this

process, the measurement model can be evaluated. All individual items in the overall

instrument do not have to perfectly align with the construct, but they should be relatively aligned across the instrument as whole (Wilson, 2005).

When reviewing fit statistics to collect validity evidence, misfit items that have fit values greater than 1.0 are important to review for validity evidence as they signal that an item contributes less to the overall estimation than other latent values or constructs. Fit values less than 1.0 within misfit items are less concerning as they usually indicate better than expected estimation of the construct (Wright & Masters, 1981). Determining misfit from mean square values is not precise as there is not an absolute limit; however, researchers have developed some criteria for more commonly accepted upper and lower bounds, 0.75 and 1.33 respectively, have been found to be useful in helping to determine misfit (Adams & Khoo, 1996; Wilson, 2005).

Based on results and analysis of the t statistics and the weighted mean square fit, it appears that the items in the SEAM instrument fit the partial credit model well. One item, AM 8, italicized in Table 18, had a mean square value of 2.85, which is significantly higher than the 1.0 level and its upper and lower limit confidence levels of .13 and 1.87, fall outside of the accepted appropriate levels. This suggests that this item was a poor fit in relation to the other items in the instrument. The rest of the item's mean square values were closer to the 1.0 level or slightly below the 1.0 level and had confidence levels within or close to the accepted parameters. One potential cause for this misfit of this particular item was that it was the only question in the survey that was framed negatively as opposed to all other items that were framed positively.

IRT analysis. IRT item analysis provided the last form of internal structure validity evidence collected in this study. The general results of the item analysis suggest that individual items across the SEAM instrument were generally stable and consistent (see Appendices K & L). The data tables below will demonstrate an example of an instrument item that did not fit, as well as two examples of items that did fit. Each data table presented provides the number of response categories for each item, the count of respondents (count) for each item, the percentage of students who answered in each response category, and the point bi-serial correlation for each response category. Response category 0 represents students that strongly disagreed, disagreed or were neutral; response category 1 represents students that agreed; and response category 2 represented students that strongly agreed.

Item analysis in this section will review the mean person location in each of the response categories. In general, we expect that students registering higher on the SEAM scale will score higher on each item (Wilson, 2005). Through a review of the mean location values, items that do not show consistency within the SEAM instrument can be identified; 19 of the 21 items displayed well-functioning mean locations that increased as response categories increased. Two items, AM 8 and SE 4 both, displayed lack of consistency.

The item analysis for AM 8, shown in Table 19, indicates that the mean person locations increase very slightly from -0.18 to -0.4 from category 0 to category 2; however, there were no respondents in category 1. This strongly suggests that AM 8 did not capture a meaningful range of response.

Table 19

Item Statistics for AM8

Statistics	Response Categories		
	0	1	2
Count	42	0	2
Percent	95.45	0	4.55
Pt-Biserial	0.1		-0.1
Mean Location	-0.18		-0.4
Std. Dev. Of Locations	1.47		1.5

Item SE 4 also demonstrates issues of consistency as demonstrated in Table 20. In SE 4 we can see responses spread across all three response categories: 8 students in category 0, 20 students in category 1, and 16 students in category 2. However, their mean locations on the SEAM scale do not correspondingly increase across each response category as expected.

Table 20

Item Statistics for SE4

Statistics	Response Categories		
	0	1	2
Count	8	20	16
Percent	18.18	45.45	36.36
Pt-Biserial	-0.28	-0.36	0.6
Mean Location	-0.76	-0.82	0.87
Std. Dev. Of Locations	1.47	0.99	1.38

As Table 20 shows, mean locations in the zero-response category start at the -0.76 logit level, decrease to the -0.82 logit level as the response category increases to 1 and then increases to the .87 logit level in response category 2. This suggests the item SE4 is not aligning with the broader SEAM instrument.

Conversely, the remaining 19 instrument items follow a similar pattern, as represented by item AM6. As Table 21 demonstrates, AM6 had relatively even distribution of student respondents across all three response categories: 13 in category 0, 19 in category 1, and 12 in category 2. The mean location of student respondents on the SEAM scale within the 0 category was at the -1.58 logit level, within the 1 category they increased to the -.01 logit level, and within the 2 category they increased to the 1.04 logit level.

Table 21

Item Statistics for AM6

Statistics	Response Categories		
	0	1	2
Count	13	19	12
Percent	29.55	43.18	27.27
Pt-Biserial	-0.57	0.09	0.48
Mean Location	-1.58	-0.01	1.04
Std. Dev. Of Locations	1.22	0.84	1.2

This demonstrates that item AM6 was successful in capturing a range of responses across student response categories that corresponded with increasing SEAM ratings.

Reliability Evidence

Internal-consistency coefficients are one of the three recognized broadly categorized methods of exploring reliability and it involves the "the relationships/interactions among scores derived from individual items or subsets of the items within a test, all data accruing from a single administration" (AERA et al., 2014, p. 37). This study generated Cronbach's alpha and person separation coefficients to document the reliability of the instrument. It is recognized that having multiple reliability coefficients can be useful as

they have different potential sources of error and can convey different angles of information (AERA et al., 2014).

In classical test theory, Cronbach's alpha is valued for its usefulness of measuring internal consistency, particularly with polytomous response data, as this study used (Cronbach, 1951). Rasch IRT modeling has its own indicator, person separation reliability (Wright & Masters, 1981). The two values are both formed from the ratio of variance between person estimates and total variance including error. In instruments where persons and items are well aligned, the values of will be close to each other as well. Table 22 contains the Cronbach's alpha and person separation values of the SEAM scale.

Table 22

Internal Consistency Reliability Coefficients

| Scale | Coefficient Values (r) | |
	Cronbach's Alpha	Person Separation
SEAM	0.91	0.94

In the absence of any clear and well-defined standards for an acceptable level of reliability evidence, any reliability coefficient above .9 is considered strong evidence of an instrument's consistency. The results of both coefficient values indicate that the SEAM instrument is reliable indicator of student's beliefs.

Summary of Research Question 3

During the course of the study, three primary forms of validity evidence were collected to validate the SEAM instrument: content validity evidence, response processes

validity evidence, and internal structure validity evidence. Relations to external variability evidence and consequence evidence was not collected. Evidence for content validity, response processes, and internal structure were presented. There are four types of reliability evidence that can be typically collected: internal consistency, alternative forms, test-retest, and rater reliability. Only one form of reliability evidence was collected for the SEAM instrument, internal consistency. Two others could have been collected, test-retest and alternate forms; this will be discussed more in Chapter V. Two different reliability coefficients were collected, Cronbach's alpha and person separation, and each showed high levels of reliability in the instrument. With the validity and reliability evidence presented it is the proposition of this study that the SEAM instrument could be used in a limited role to measure academic mindset and self-efficacy constructs in students.

Chapter V: Discussion

Summary of the Study

There is much research to show the potential impact of noncognitive factors on student achievement, but there is still much research to be done to understand the level of interplay between the various noncognitive factors and their impact on student achievement (Bandura & Schunk, 1981; Bandura, 1986; Bouffard-Bouchard, 1990; Cury et al., 2006; Dweck & Leggett, 1988; Farrington et al., 2012; Lent et al., 1984; Pajares, 1996; Schunk & Hanson, 1985). Changes to the California school accountability framework, that have required schools and districts to collect a wider range of school climate and student engagement data, have created opportunities to study the level of direct impact noncognitive factors may be having on student achievement. Collecting and using sensitive data to help drive decision making for school planning and budgeting purposes comes with important responsibilities. There is a responsibility to understand, to the best of our ability, the roles that noncognitive factors have on student achievement; this will help decision makers and stakeholders make informed decisions about how much time and money should be spent on developing noncognitive factors. Educational decision makers also have a responsibility to continue to study and understand the full range of factors that relate to student achievement in order to ensure that there are no other priorities or interventions that may have more critical role. Lastly, educational decision makers have an obligation to responsibly and ethically understand and use data collection instruments; if they do not, they risk misrepresenting or misunderstanding what is going on in their schools, or even harming students.

The purpose of this study was to explore the relationship between academic mindset and self-efficacy, and student achievement, as well as to explore this relationship in relation to other factors that may also impact student achievement. This study explored these relationships through the creation of the SEAM survey, which was comprised of modifications made to the MESH survey, a survey widely used in the state of California to measure noncognitive levels of students. In creating the SEAM survey, robust validation and reliability protocols informed by leaders in this field were used to create an efficacious instrument to detect these noncognitive factors in students (Pellegrino et al., 2001; Wilson, 2005) The purpose of this work was to advance the cause of further integrating noncognitive factor measurement with student achievement measurement. In the following sections are implications for each of the three research questions and recommendations that arose from this research.

Summary of RQ1: Implications

The first research question was: "What is the relationship, if any, between noncognitive factors and mathematics achievement for at promise middle school students?" This question was further broken down into two sub parts: "What is the relationship, if any, between self-efficacy (SE) and mathematics achievement for at promise middle school students?" and, "What is the relationship, if any, between academic mindset (AM) and mathematics achievement for at promise middle school students?" This question was designed to study the relationship between self-efficacy and academic mindset, and student math achievement. Evidence from the correlational analysis demonstrated that the SEAM survey detected statistically significant positive

110

correlations between 4 of the academic mindset instruments (AM1, 3, 5 & 11) and math achievement, as well as 1 of the self-efficacy instrument items (SE9) and math achievement. Four of the other academic mindset items and ten of the self-efficacy items also showed statistically insignificant positive correlations. Similarly, when the self-efficacy and academic mindset items were combined during the IRT analysis process into the overall SEAM scale score and correlated to math achievement, only modest correlation was detected. This would suggest that academic mindset and self-efficacy may have some moderate impact on student achievement, as measured by RIT.

Summary of RQ2: Implications

The second research question was: "What factors predict math achievement?" This question was designed to explore demographic factors that may also be impacting student achievement. The study findings strongly suggest that student socio-economic status has a high impact on student achievement, as measured by standardized test scores. Students of lower socio-economic status, as measured by free and reduced lunch status, were less likely to outperform their peers of higher socio-economic status. The study also found evidence that self-efficacy and academic mindset beliefs, as measured through the SEAM instrument, had a moderate and statistically significant relationship with math achievement as measured by RIT. This suggests that students' perceptions of their academic mindset and self-efficacy are some of the drivers of student academic achievement. Student's socio-economic status is not something that can be quickly or easily changed by educators; however, helping students to understand and develop academic mindset or self-efficacy is something that has been demonstrated to be possible

and can even be done in relatively short periods of time (Yeager et al., 2019). This study's findings suggest that these types of foci or interventions might have some measurable and cost-effective payback if deployed with consideration of local contexts and with realistic expectations of results.

Summary of RQ3: Implications

The third research question was: "How reliably does the SEAM indicator measure levels of self-efficacy and academic mindset and what is the evidence for the validity of SEAM scores to measure students?" The purpose of this research question was to ensure that high standards were used in the creation of the SEAM instrument and to collect validity and reliability evidence that would support further usage of the instrument. This study was able to collect three of the five types of validity evidence commonly accepted in validation work: content validity evidence, response processes validity evidence, and internal structure validity evidence.

The study's content validity evidence was generated through two construct maps, one for academic mindset and the other for self-efficacy. These construct maps were grounded in the literature and reviewed by expert panels. Response process evidence was collected from two think-out-loud protocols run with two different middle school students; their feedback led to revisions and further iteration of the instrument to increase alignment of the instrument to the construct. Lastly, internal structure evidence was generated through the development of a Wright Map, model fit analysis, and IRT analysis. Evidence collected from these sources showed that the SEAM instrument was calibrated to capture a range of self-efficacy and academic mindset beliefs and that the

SEAM instrument captured a range of responses on the logit scale. This validity evidence was combined with the fit analysis data, which showed that all of the items, with the exception of AM 8, were generally aligned with the SEAM model. This relatively strong alignment provides more validity evidence for the instrument. Lastly, IRT analysis provided the third leg of internal structure validity evidence. Evidence collected in this study showed that only two of the twenty-two items in the instrument were out of alignment and stability with each other. This suggests relatively strong alignment in the construction of the SEAM instrument.

The study was also able to collect reliability evidence from one of the four widely recognized forms of validity evidence, internal consistency. There were two forms of reliability evidence, test-retest and alternate forms, that could have also been collected but were not. This study generated two measures of internal consistency evidence in the form of a Cronbach's alpha value and a person separation value. Both displayed strong evidence for the internal consistency and reliability of the SEAM instrument.

Discussion

The current school accountability model in California, LCAP, was designed to foster continuous improvement by giving local agencies and schools the power to set their own goals, plan actions to reach those goals, and harness resources to meet and evaluate those goals. This school accountability model gives schools and local school agencies greater flexibility to set priorities and measures for success beyond the previous accountability framework's focus on standardized achievement, but schools and districts are still learning to use this greater flexibility to focus on the right drivers for continuous

improvement (Fullan & Rincon-Gallardo, 2017). This new accountability system requires school plans, budgets, and improvement efforts to not only focus on academic achievement as a measure for school progress, but also the development of the whole child. Along with greater emphasis on whole-child education and support has come greater focus on school climate and student engagement. These are factors that have less to do with classical cognitive abilities, such as intelligence, and more to do with beliefs and attitudes. Schools are now required to study these beliefs and attitudes in their students and plan continuous improvement efforts in relation to them, as well as academic achievement. With this imperative outlined in the LCAP system, there is a greater need to measure and quantify these noncognitive factors and their impact on student achievement.

Noncognitive Factors and Student Achievement

Educators and researchers have developed and identified many noncognitive constructs that they believe may have interconnected relationships on student achievement outside classical notions of cognitive factors such as intelligence. In education, some of these noncognitive factors, like academic mindset and self-efficacy, are prevalent in teacher training literature and professional development. However, despite voluminous and rich research into these noncognitive factors, there is much we still don't know and the leading experts in this field will admit that there is still much we do not understand about how these noncognitive factors interact with one another and how much they really directly impact student achievement (Farrington et al., 2012). School districts and schools are spending millions of dollars on professional

development, trainings, instruments, and curriculum to support the development of these noncognitive factors in their students and staff but we still are trying to understand how these noncognitive factors work.

LCAP has put more impetus on school and district plans, budgets, and foci to improve student learning outcomes and develop the whole child. This has created an imperative for researchers and educators to robustly and systematically study how these noncognitive factors are interacting with cognitive factors and student achievement. Educators, researchers, and policy makers have a responsibility to study and attempt to understand how these various factors interact with each other to ensure to the best of our ability that our interventions, resources, and professional development are targeted and meaningful.

There is still work to be done in the area of developing the systemic impact of noncognitive factors. We must ensure that the noncognitive factors schools are trying to understand from school climate data are not just minor notes in a school plan, but rather something schools are actively seeking to cultivate and inspire. In my own experience, I have seen how academic data and school climate data can be combined and studied to provide a rich tapestry of what is going on in schools and school systems and can be used to help improve student learning outcomes. I have also seen how this data can be collected simply for the sake of reporting out in accreditation reports, school plans, and budgets. The findings of this study align with previous noncognitive factor research studies that show moderate connections between noncognitive factors like self-efficacy and academic mindset and student achievement. This suggests that a moderate amount of

investment and professional development in these areas would be prudent if there is evidence that these noncognitive factors were malleable.

Noncognitive Factors Promise and Limitations

Various research has shown that noncognitive factors are malleable and open to development and improvement (Blackwell et al., 2007; Farrington et al., 2012; Yeager & Walton, 2011). There is evidence to suggest that interventions and programs designed to develop noncognitive factors can be effective. There is also evidence that these noncognitive factors can and do impact student achievement. Therefore, researchers and educators must continue to study this relationship and look for interventions, professional development, and programs that can help develop noncognitive factors without allowing these sorts of programs or interventions to take over the curricular or school programs. Research has also shown that some relatively low cost and minimally disruptive interventions can have moderate impacts on developing noncognitive factors, but the long-lasting nature of these interventions is still being studied and understood (Yeager et al., 2019). This study explored the relationship between various demographic factors (English learner, gender, socio-economic status, and ethnicity) as well as their levels of self-efficacy and academic mindset to see which of these factors had any impact on student achievement, as measured by student MAP Growth scores. The evidence of this study suggests that students with higher ratings of self-efficacy and academic mindset, as measured by the SEAM instrument, were more likely to score higher on the MAP Growth assessment. However, there was evidence found in this study that there was

another factor that had an even greater likelihood of predicting MAP scores: socio-economic status.

This study provided evidence, aligned with previous research, that suggests one of the greatest predictors of student achievement (when measured by standardized tests) is socio-economic status (Berliner, 2013, 2014). The combined findings brought forth in this study, of the relationship between socio-economic status and math achievement, as well as between student's self-efficacy and academic mindset beliefs and math achievement, raise questions to wrestle within our educational system. There is strong evidence that systemic income inequality and poverty are contributing factors in chronic under-performance on standardized assessments in school populations across the state of California and the nation. There is also evidence that noncognitive factors may have an impact on these standardized assessment scores. This evidence has drawn researchers and educators into the practice of developing and using instruments and interventions designed to track, monitor, and improve these noncognitive factors. Since there is evidence to suggest that this work could be beneficial if done in a cost effective and minimally disruptive manner, it is the recommendation of this study that this work should continue to be done.

Other research has demonstrated the potential for interventions designed to develop growth mindset and prevent gender stereotype differences in mathematics achievement (Lee et al., 2021). While this study's limited sample did not detect differences in achievement related to gender, the phenomenon has been documented in other studies and it is an area of concern (Else-Quest et al., 2010; Fredricks & Eccles, 2002; Weidinger

et al., 2018). In future research, the SEAM survey could be improved by studying gender-based differences in responses and perceptions of self-efficacy and academic mindset.

The instruments used to study, track, and understand the noncognitive factors and their interplay with academic outcomes must be created with robust and thorough standards in order to ensure their validity and reliability. This study followed the guiding principles and standards of validity and reliability outlined by the NRC Assessment Triangle and Mark Wilson's Four Building Blocks of instrument development (Pellegrino et al., 2001; Wilson, 2005). This study collected three types of validity evidence: content validity, response process validity, and internal structure validity evidence. There were also two forms of validity evidence that were not collected but would have been possible to collect and would have been valuable: external variable validity evidence and consequential validity evidence.

This study could certainly benefit from follow up research studying how the SEAM survey relates to other noncognitive factor surveys. It would also be interesting and valuable to study how the SEAM survey might be related to other academic indicators, such as grades or other test scores, in order to collect other forms of validity evidence. In order to responsibly use the SEAM instrument in future studies, consequential validity evidence needs to be collected in follow up research as well; a researcher cannot ethically use and make policy from an instrument without understanding the social consequences of the instrument and its purpose first (Duckor, 2017; Maul et al., 2018).

This study collected one of the four types of reliability evidence recognized by the NRC Assessment Triangle and Wilson's Four Building Blocks: internal consistency. The Cronbach's Alpha and Person Separation coefficient evidence provided in this study both suggested that the SEAM instrument was a stable and reliable instrument. However, further reliability evidence could have been collected to make an even stronger case for the reliability of the instrument. Future studies could create alternate forms of the SEAM instrument in order to see if the same student would score similarly on different versions of the instrument. Also, test-retest protocols could be used to see if the student consistently scores similarly when using the same instrument in close succession (Wilson, 2005). Both of these additional forms of reliability evidence could be pursued further in follow up research to provide further reliability evidence of the SEAM instrument.

Validation is an ongoing process and a matter of degree; when using any instrument, the reliability and validation should be checked for each additional subgroup tested (Messick, 1989). If educators are going to continue researching how noncognitive factors and academic achievement are interrelated then we must continue to refine, validate and ensure the reliability of whatever instrumentation we use.

Study Limitations

In addition to the validity and reliability checks that could make the SEAM instruments use more sound, there are a few other key study limitations. Attributional bias can be a factor in any survey that asks respondents to self-report on feelings or beliefs and there is an opportunity to attribute those feelings or beliefs positively or in a

self-serving way. It is a documented phenomenon that people are more likely to make positively biased attributions that self-serve than negative attributions (Mezulis et al., 2004). It is possible that 8th grade students completing a survey may be more likely to answer the way they think an educator or adult wants them to answer, even if their answers are anonymous and they do not know the researcher. A mixed method study with an opportunity to follow up with students more qualitatively could have assisted in teasing this out further.

The study is also limited by the conclusions that can be drawn from correlational and regression analysis. There are several well-documented limitations of these types of analyses: researchers cannot know for certain that the independent variable studied is definitively affecting the dependent variable, these types of studies are prone to manipulation of outliers which dramatically impact results, and they are more stable with larger amounts of data (Spiegler, 2019) It is certainly likely that there are factors impacting the student's mathematics achievement other than factors studied. All students who completed the survey were included in all statistical analysis provided, no outliers were removed or manipulated in this study.

Quantitative studies gain much of their strength from larger numbers of respondents. While much care was put into the development of a validated and reliable instrument, the smaller sample size of this study makes it difficult to make widely generalizable claims from the findings of this study. While a large pool of parents was solicited for permission to work with their students, relatively few responded due to circumstances beyond my control.

Students were sampled in the fall of 2020, during a global pandemic COVID-19, widespread fires across the county, and major social and economic dislocations which forced the district to move to fully virtual instruction. The combination of the timing of this study, which necessarily coincided with the beginning of the year testing, the steep learning curve of virtual instruction, and overwhelmed students and parents during the adjustment, impacted parent responses and student participation in this study.

Conclusion

Despite the challenges and obstacles of conducting a study during a global pandemic, this study was able to explore the connections between the noncognitive factors of self-efficacy, academic mindset, and math achievement. In addition, the study was also able to explore how other demographic factors may or may not relate to math achievement. During the course of this study, a new instrument, the SEAM survey, was developed from the MESH survey; that SEAM survey was used to study the relationship of noncognitive factors and math achievement. Findings support previous research that there are connections between noncognitive factors and achievement and that certain demographic factors play a large role in shaping achievement, as measured by standardized testing. Through the process of developing the instrument, the researcher was able to practice valuable reliability and validation evidence collecting techniques.

This study was important because it invited study and exploration into the relationship of noncognitive factors and achievement. This is important work in the state of California, where whole-child support is built into the continuous improvement accountability framework of the state. The challenge is that there is still much to

understand about this relationship, and we need educators to continue to be trained and supported in their efforts to build school systems that can sustain this improvement. This study advanced three recommendations. First, continue to study the relationship between noncognitive factors and achievement and develop instruments that measure this relationship in order to aid school improvement efforts. Second, explore, study, and develop cost effective and minimally disruptive interventions that do not take over the curriculum and meaningful learning experiences of the students, but do help them develop noncognitive factors that have been shown to impact academic achievement. And lastly, look for, and expect, rigorous and robust validity and reliability checks on any instruments used in the two aforementioned recommendations, to ensure that vital policy decisions and school planning are accomplished with the best information possible.

References

Adams, R. A., & Khoo, S.-T. (1996). *ACER Quest: The interactive test analysis system*. Australian Council for Educational Research (ACER). https://research.acer.edu.au/measurement/3/

American Educational Research Association (AERA), National Council on Measurement in Education, & American Psychological Association. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

American Institute of Research. (2016). *Screening tools for chart rating system*. National Center on Intensive Intervention. https://charts.intensiveintervention.org/ascreening

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260-267. https://doi.org/10.1037/0022-0663.80.3.260

Amrein-Beardsley, A., & Collins, C. (2012). The SAS education value-added assessment system (SAS® EVAAS®) in the Houston independent school district (HISD): Intended and unintended consequences. *Education Policy Analysis Archives, 20*. https://doi.org/10.14507/epaa.v20n12.2012

Baker, E. L. (1997). Model-based performance assessment. *Theory into Practice, 36*(4), 247-254. https://doi.org/10.1080/00405849709543775

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist, 33*(4), 344-358. https://doi.org/10.1037/0003-066X.33.4.344

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W.H. Freeman and Company.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1-26. https://doi.org/10.1146/annurev.psych.52.1.1

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science, 1*(2), 164-180. https://doi.org/10.1111/j.1745-6916.2006.00011.x

Bandura, A. (2008). *Reconstrual of "free will" from the agentic perspective of social cognitive theory*. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (p. 86–127). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195189636.003.0006

Bandura, A. (2012). *Social cognitive theory*. In P. A. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 1, pp. 349-374). SAGE Publications Ltd. https://www.doi.org/10.4135/9781446249215.n18

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology, 45*(5), 1017-1028. https://doi.org/10.1037/0022-3514.45.5.1017

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

Berliner, D. (2013). Effects of inequality and poverty vs. teachers and schooling on america's youth. *Teachers College Record, 115*(12). https://eric.ed.gov/?id=EJ1020258

Berliner, D. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record, 116*(1). https://eric.ed.gov/?id=EJ1020233

Berliner, D. (2018). Between scylla and charybdis: Reflections on and problems associated with the evaluation of teachers in an era of metrification. *Education Policy Analysis Archives, 26* https://doi.org/10.14507/epaa.26.3820

Berliner, D. C., & Biddle, B. J. (1996a). The manufactured crisis: Myths, fraud, and the attack on America's public schools. Basic Books.

Berliner, D. C., & Biddle, B. J. (1996b). The manufactured crisis: Myths, fraud, and the attack on America's public schools. *NASSP Bulletin, 80*(576), 119-121. https://doi.org/10.1177/019263659608057619

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

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*(1), 246-263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Boaler, J. (2016). In C. Dweck (Ed.), *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages, and innovative teaching*. Jossey-Bass & Pfeiffer Imprints.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Weel, B. T. (2008). The economics and psychology of personality traits. *Journal of Human Resources, 43*(4), 972-1059. https://doi.org/10.1353/jhr.2008.0017

Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *The Journal of Social Psychology, 130*(3), 353-363. https://doi.org/10.1080/00224545.1990.9924591

Bouffard-Bouchard, T., Parent, S., & Larivee, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. *International Journal of Behavioral Development, 14*(2), 153-164. https://doi.org/10.1177/016502549101400203

Bowles, S. (1976). In H. Gintis (Ed.), *Schooling in capitalist America: Educational reform and the contradictions of economic life*. Basic Books.

Bridgeland, J. M., Diiulio, J. J., & Morison, K. B. (2006). *The silent epidemic: Perspectives of high school dropouts* [Report]. Civic Enterprises; Peter D. Hart Research Associates. https://files.eric.ed.gov/fulltext/ED513444.pdf

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology, 79*(4), 474-482. https://doi.org/10.1037/0022-0663.79.4.474

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology, 58*(1), 1-14. https://doi.org/10.1111/j.2044-8279.1988.tb00874.x

California Department of Education. (2020). English language arts/literacy and mathematics: Smarter Balanced summative assessments [2018-2019 Data]. https://caaspp-elpac.cde.ca.gov/caaspp/DashViewReport?ps=true&lstTestYear=2019&lstTestType=B&lstGroup=1&lstSubGroup=1&lstSchoolType=A&lstGrade=13&lstCounty=00&lstDistrict=00000&lstSchool=0000000

Caprara, G. V., Barbaranelli, C., Borgogni, L., & Steca, P. (2003). Efficacy beliefs as determinants of teachers' job satisfaction. *Journal of Educational Psychology, 95*(4), 821-832. https://doi.org/10.1037/0022-0663.95.4.821

Choi, N., Fuqua, D. R., & Griffin, B. W. (2001). Exploratory analysis of the structure of scores from the multidimensional scales of perceived self-efficacy. *Educational and Psychological Measurement, 61*(3), 475-489. https://doi.org/10.1177/00131640121971338

Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences of the United States of America, 113*(31), 8664. https://doi.org/10.1073/pnas.1608207113

Conley, D. (2007). *Toward A more comprehensive conception of college readiness.* Educational Policy Improvement Center. https://docs.gatesfoundation.org/documents/collegereadinesspaper.pdf

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/10.1007/BF02310555

Cury, F., Elliot, A. J., Fonseca, D. D., & Moller, A. C. (2006). The social-cognitive model of achievement motivation and the 2×2 achievement goal framework. *Journal of Personality and Social Psychology, 90*(4), 666-679. https://doi.org/10.1037/0022-3514.90.4.666

Darling-Hammond, L., Bae, S., Cook-Harvey, C., Lam, L., Mercer, C., Podolsky, A., Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act* [Report]. Learning Policy Institute; Stanford Center for Opportunity Policy in Education. https://learningpolicyinstitute.org/product/pathways-new-accountability-through-every-student-succeeds-act

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (Eds.). (1999). *How people learn: Bridging research and practice.* The National Academies Press. https://doi.org/10.17226/9457

Duckor, B. (2017). Got grit? Maybe. *Phi Delta Kappan, 98*(7), 61-66. https://doi.org/10.1177/0031721717702634

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048. https://doi.org/10.1037/0003-066X.41.10.1040

Dweck, C. (2008). *Mindset: The new psychology of success* (1st ed.). Random House.

Dweck, C. S., Davidson, W., Nelson, S., & Enna, B. (1978). Sex differences in learned helplessness: II. the contingencies of evaluative feedback in the classroom and III. an experimental analysis. *Developmental Psychology, 14*(3), 268-276. https://doi.org/10.1037/0012-1649.14.3.268

Dweck, C., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273. https://doi.org/10.1037/0033-295X.95.2.256

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

Dweck, C. S., Walton, G. M., & Cohen, G. L. (2014). *Academic tenacity: Mindsets and skills that promote long-term learning.* Bill & Melinda Gates Foundation. https://ed.stanford.edu/sites/default/files/manual/dweck-walton-cohen-2014.pdf

Ehrlinger, J., Mitchum, A. L., & Dweck, C. S. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology, 63*, 94-100. https://doi.org/10.1016/j.jesp.2015.11.001

Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology, 77*(2), 162-173. https://doi.org/10.1037/0022-0663.77.2.162

Else-Quest, N., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127. https://doi.org/10.1037/a0018053

Farkas, G. (2003). Cognitive skills and noncognitive traits and behaviors in stratification processes. *Annual Review of Sociology, 29*(1), 541-562. https://doi.org/10.1146/annurev.soc.29.010202.100023

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance—A critical literature review.* The University of Chicago Consortium on Chicago School Research. https://www.kipp.org/wp-content/uploads/2016/11/Teaching_Adolescents_to_Become_Learners.pdf

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). McGraw-Hill.

Fowler, F. J. (2009). *Survey research methods* (4th ed.). SAGE Publications.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*(4), 519-533. https://doi.org/10.1037/0012-1649.38.4.519

Fullan, M., & Rincon-Gallardo, S. (2017). *California's golden opportunity taking stock: Leadership from the middle.* Motion Leadership. https://michaelfullan.ca/wp-content/uploads/2017/09/17_Californias-Golden-Opportunity-Taking-Stock-FinalAug31.pdf

Glass, G. V. (2008). *Fertilizers, pills, and magnetic strips: The fate of public education in America.* Information Age Publishing.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*(6), 645-662. https://doi.org/10.1016/j.appdev.2003.09.002

Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85*(3), 541-553. https://doi.org/10.1037/0022-3514.85.3.541

Gross, G. (2013, July 31). *The value of standardized testing*. Huffington Post. https://www.huffpost.com/entry/the-value-of-standardized_b_3684212

Hattie, J. (2018a). *Collective teacher efficacy.* Corwin Visible Learning. http://www.visiblelearningmetax.com/Influences/View/collective_teacher_efficacy

Hattie, J. (2018b). *Growth vs. fixed mindsets.* Corwin Visible Learning. http://www.visiblelearningmetax.com/influences/view/growth_vs._fixed_mindsets

Hattie, J. (2018c). *Self-efficacy.* Corwin Visible Learning. http://www.visiblelearningmetax.com/influences/view/self-efficacy

Heckman, J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review, 91*(2), 145-149. https://doi.org/10.1257/aer.91.2.145

Heckman, J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411-482. https://doi.org/10.1086/504455

Herman, J. L., & Haertel, E. (Eds.) (2005). *Uses and misuses of data for educational accountability and improvement*. Blackwell Publishing.

Jencks, C. (1979). *Who gets ahead? The determinants of economic success in America*. Basic Books.

Kane, M. B., & Mitchell, R. (Eds.). (1996). *Implementing performance assessment: Promises, problems, and challenges.* Lawrence Erlbaum Associates.

Kim, C., Park, S. W., Cozart, J., & Lee, H. (2015). From motivation to engagement: the role of effort regulation of virtual high school students in mathematics courses. *Educational Technology & Society*, *18*(4), 261+. https://psycnet.apa.org/record/2015-47277-019

Kohn, A. (2011). The case against grades. *Educational Leadership, 69*(3), 28-33. https://www.alfiekohn.org/article/case-grades/

Larocca, R., & Krachman, S. B. (2016). *Expanding the definition of student success under ESSA: Opportunities to advance social-emotional mindsets, skills, and habits for today's students* [Policy brief]. Transforming Education. https://files.eric.ed.gov/fulltext/ED605420.pdf

Lee, J., Lee, H. J., Song, J., & Bong, M. (2021). Enhancing children's math motivation with a joint intervention on mindset and gender stereotypes. *Learning and Instruction, 73*, 101416. https://doi.org/10.1016/j.learninstruc.2020.101416

Lent, R. W., Brown, S. D., & Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. *Journal of Counseling Psychology, 31*(3), 356-362. https://doi.org/10.1037/0022-0167.31.3.356

Linn, R. L. (1988). *Educational measurement* (3rd ed.). Collier Macmillan Publishers.

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research, 37*(3), 888-902. https://doi.org/10.1016/j.ssresearch.2008.03.004

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley Publishing Company.

Maul, A., Mari, L., Irribarra, D. T., & Wilson, M. (2018). The quality of measurement results in terms of the structural features of the measurement process. *Measurement: Journal of the International Measurement Confederation, 116*, 611-620. https://doi.org/10.1016/j.measurement.2017.08.046

McDermott, R. (1993). The acquisition of a child by learning a learning disability. In S. Chaiklin, & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 269-305). Cambridge University Press.

Mendoza-Denton, R., Goldman-Flythe, M., Pietrzak, J., Downey, G., & Aceves, M. J. (2010). Group-value ambiguity: Understanding the effects of academic feedback on minority students' self-esteem. *Social Psychological and Personality Science, 1*(2), 127-135. https://doi.org/10.1177/1948550609357796

Mendoza-Denton, R., Pietrzak, J., & Downey, G. (2008). Distinguishing institutional identification from academic goal pursuit: Interactive effects of ethnic identification and race-based rejection sensitivity. *Journal of Personality and Social Psychology, 95*(2), 338-351. https://doi.org/10.1037/0022-3514.95.2.338

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement, 21*(3), 215-237. https://doi.org/10.1111/j.1745-3984.1984.tb01030.x

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11. https://doi.org/10.3102/0013189X018002005

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*(5), 711-747. https://doi.org/10.1037/0033-2909.130.5.711

Midgley, C., Anderman, E., & Hicks, L. (1995). Differences between elementary and middle school teachers and students: A goal theory approach. *The Journal of Early Adolescence, 15*(1), 90-113. https://doi.org/10.1177/0272431695015001006

Mislevy, R. J. (1994). Test theory reconceived. *ETS Research Report Series, 1994*(1), i-38. https://doi.org/10.1002/j.2333-8504.1994.tb01575.x

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38. https://doi.org/10.1037/0022-0167.38.1.30

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* https://edreform.com/wp-content/uploads/2013/02/A_Nation_At_Risk_1983.pdf

Nichols, P. (1994). A framework for developing cognitively diagnostic assessments. *Review of Educational Research, 64*(4), 575-603. https://doi.org/10.3102/00346543064004575

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Harvard Education Press.

Niehaus, K., Rudasill, K. M., & Adelson, J. L. (2012). Self-efficacy, intrinsic motivation, and academic outcomes among Latino middle school students participating in an after-school program. *Hispanic Journal of Behavioral Sciences, 34*(1), 118-136. https://doi.org/10.1177/0739986311424275

Northwest Evaluation Association. (2017). *Linking the smarter balanced assessments to NWEA MAP growth tests*. https://lasierra.edu/fileadmin/documents/education/curriculum-and-instruction/MAP/teachers-files/sbac-map-growth-linking-study.pdf

Northwest Evaluation Association. (2019). *MAP growth technical report*. [Internal Proprietary Technical Report Provided to Researcher]

O'Rourke, E., Haimovitz, K., Ballweber, C., Dweck, C., & Popović, Z. (2014, April 26). *Brain points: A growth mindset incentive structure boosts persistence in an educational game* [Paper presentation]. Association for Computing Machinery's SIGCHI Conference on Human Factors in Computing Systems (CHI '14), New York, NY. https://doi.org/10.1145/2556288.2557157

Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology, 91*(1), 188-204. https://doi.org/10.1037/0022-3514.91.1.188

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578. https://doi.org/10.3102/00346543066004543

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24*(2), 124-139. https://doi.org/10.1006/ceps.1998.0991

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 20*(4), 426-443. https://doi.org/10.1006/ceps.1995.1029

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193-203. https://doi.org/10.1037/0022-0663.86.2.193

Pajares, F., & Schunk, D. (2002). *Self and self-belief in psychology and education: A historical perspective*. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (p. 3–21). Academic Press. https://doi.org/10.1016/B978-012064455-1/50004-X

Pellegrino, J. W., Baxter, G. P., & Glaser, R. (1999). Addressing the "two disciplines" problem: Linking theories of cognition and learning with assessment and instructional practice. *Review of Research in Education, 24*, 307-353. https://doi.org/10.3102/0091732X024001307

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* The National Academies Press. https://doi.org/10.17226/10019

Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (Eds.). (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress.* The National Academies Press. https://doi.org/10.17226/6296

Pietsch, J., Walker, R., & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology, 95*(3), 589-603. https://doi.org/10.1037/0022-0663.95.3.589

Pintrich, P. R. (1999). Taking control of research on volitional control: Challenges for future theory and research. *Learning and Individual Differences, 11*(3), 335-354. https://doi.org/10.1016/S1041-6080(99)80007-7

Polikoff, M. S. (2019). *Gauging the revised California school dashboard: Evidence from the 2019 PACE/USC Rossier voter poll.* Policy Analysis for California Education. https://edpolicyinca.org/sites/default/files/PB_Polikoff_Feb-2019.pdf

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology, 103*(3), 683-700. https://doi.org/10.1037/a0023911

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). University of Chicago Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173-220). Academic Press.

Samejima, F. (1977). A use of the information function in tailored testing. *Applied Psychological Measurement, 1*(2), 233-247. https://doi.org/10.1177/014662167700100209

Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement, 18*(3), 229-244. https://doi.org/10.1177/014662169401800304

Schunk, D. (1983). Reward contingencies and the development of children's skills and self-efficacy. *Journal of Educational Psychology, 75*(4), 511-518. https://doi.org/10.1037/0022-0663.75.4.511

Schunk, D. (Ed.). (1995). *Self-efficacy and education and instruction*. Plenum Press.

Schunk, D. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal, 33*(2), 359-382. https://doi.org/10.3102/00028312033002359

Schunk, D., & Hanson, A. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*(3), 313-322. https://doi.org/10.1037/0022-0663.77.3.313

Shaw, L. (2013, March 30). Educators debate validity of MAP testing. *Seattle Times.* https://www.seattletimes.com/seattle-news/educators-debate-validity-of-map-testing/

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychological Science, 29*(4), 549-571. https://doi.org/10.1177/0956797617739704

Skitka, L. J., Mullen, E., Griffin, T., Hutchinson, S., & Chamberlin, B. (2002). Dispositions, scripts, or motivated correction? Understanding ideological differences in explanations for social problems. *Journal of Personality and Social Psychology, 83*(2), 470-487. https://doi.org/10.1037/0022-3514.83.2.470

Slapik, M. (2017, October 17). The purpose of education—According to students. *The Atlantic.* https://www.theatlantic.com/education/archive/2017/10/the-purpose-of-education-according-to-students/541602/

Smiley, P. A., & Dweck, C. S. (1994). Individual differences in achievement goals among young children. *Child Development, 65*(6), 1723-1743. https://doi.org/10.1111/j.1467-8624.1994.tb00845.x

Spiegler, J. (2019). In Smith J. (Ed.), *Regression analysis for beginners: Exploring gun violence*. SAGE Publications.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *The American Psychologist, 52*(6), 613. https://doi.org/10.1037/0003-066X.52.6.613

Strauss, V. (2013, January 11). Teachers refuse to give standardized test at Seattle high schools—Update. *Washington Post.* https://www.washingtonpost.com/news/answer-sheet/wp/2013/01/11/teachers-refuse-to-give-standardized-test-at-seattle-high-school/

Target District. (2019). *2018-2019 NWEA MAP results.* [Internal Testing Data From Target District]

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments: Synthesis report 44.* National Center on Educational Outcomes. https://nceo.umn.edu/docs/OnlinePubs/Synth44.pdf

Transforming Education . (2016). *Measuring MESH: Student and teacher surveys curated for the CORE districts.* https://transformingeducation.org/wp-content/uploads/2017/04/160406_MeasuringMESH_ForRelease2.pdf

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783-805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202-248. https://doi.org/10.3102/00346543068002202

Valentine, J. C., Dubois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133. https://doi.org/10.1207/s15326985ep3902_3

Walton, G., & Cohen, G. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*(1), 82-96. https://doi.org/10.1037/0022-3514.92.1.82

Walton, G., & Spencer, S. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*(9), 1132-1139. https://doi.org/10.1111/j.1467-9280.2009.02417.x

Webb. (1999). *Alignment of science and mathematics standards and assessments in four states. research monograph No. 18*. ERIC - Education Resources Information Center. https://eric.ed.gov/?id=ED440852

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Development, 89*(2), e138-e156. https://doi.org/10.1111/cdev.12806

Weiss, David J. (2004). Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education. *Measurement and Evaluation in Counseling and Development, 37*(2), 70–84. https://doi.org/10.1080/07481756.2004.11909751

Wellman, H. M. (1992). *The child's theory of mind*. MIT Press.

Wentzel, K. (1998). Social relationships and motivation in middle school: The role of parents, teachers, and peers. *Journal of Educational Psychology, 90*(2), 202-209. https://doi.org/10.1037/0022-0663.90.2.202

West, M. R., & Schwerdt, G. (2012). The middle school plunge: Achievement tumbles when young students change schools. *Education Next, 12*(2). https://www.educationnext.org/the-middle-school-plunge/

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Routledge. https://doi.org/10.4324/9781410611697

Wilson, T. D., & Linville, P. W. (1985). Improving the performance of college freshmen with attributional techniques. *Journal of Personality and Social Psychology, 49*(1), 287-293. https://doi.org/10.1037/0022-3514.49.1.287

Wirt, M. W., & Wirt, F. M. (Eds.). (2009). *The political dynamics of American education* (4th ed.). McCutchan Publishing Corporation.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163-183. https://doi.org/10.1207/s15324818ame1802_2

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test, motivation, and mentally taxing items. *Applied Measurement in Education, 8*(4), 341-351. https://doi.org/10.1207/s15324818ame0804_4

Wright, B., & Masters, G. (1981). *Rating scale analysis*. MESA Press.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho…Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*(7774), 364. https://doi.org/10.1038/s41586-019-1466-y

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

Yeh, S. S. (2006). Reforming federal testing policy to support teaching and learning. *Educational Policy, 20*(3), 495-524. https://doi.org/10.1177/0895904805284119

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17. https://doi.org/10.1207/s15326985ep2501_2

Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series, 2012*(1), i-30. https://doi.org/10.1002/j.2333-8504.2012.tb02290.x

Appendix A: Parent Permission Form for Survey

REQUEST FOR YOUR PARTICIPATION IN RESEARCH

THE ROLE OF SELF-EFFICACY AND GROWTH MINDSET ON MIDDLE SCHOOL MATH
ACHIEVEMENT FOR AT PROMISE YOUTH: AN EXPLANATORY STUDY
Michael Mansfield, Education Leadership Doctoral Student & Brent Duckor, Ph.D. Faculty Supervisor

San Jose State University

PURPOSE

The purpose of this study is to examine the relationship between student beliefs about what they are
capable of doing (self-efficacy), their beliefs about whether or not they can grow in their ability (growth
mindset) and their math achievement.

PROCEDURES

Your student will be invited to participate in a brief online survey that will take place during their math
class. The survey will consist of questions pertaining to their beliefs about their academic capabilities and
about their ability to grow academically. These questions will be asked after they have completed the
annual fall administration of the NWEA MAP assessment. The survey should take 15-25 minutes to
complete. All reported information will be confidential. No personally identifying information will be
reported in the final study. Where data is reported, it will be reported in combination with other data with
no individually identifiable information included. The survey will be given during the testing period, if you
or your student chooses not to participate they will do the regularly assigned quiet activity directed by the
teacher instead of the survey.

POTENTIAL RISKS

There are minimal risks to your student associated with this study. This study involves a test that is
typically administered to all students in the district and survey questions about their academic abilities and
belief in their ability to grow academically, these questions would be no different than typical questions a
teacher may ask of a student during normal classroom activity. However, these survey questions, asked
after taking the test, may make them feel a little uncomfortable, especially if they did not do well on the test
and that made them angry or sad.

POTENTIAL BENEFITS

There are not any anticipated direct benefits to students who participate in the study. The information that
is provided may help researchers better understand the usefulness of the survey questions used in the
survey and better understand the relationship of student beliefs about their academic abilities, student
beliefs about their ability to grow academically and math achievement.

COMPENSATION

No compensation will be given for participating in this study.

CONFIDENTIALITY

Survey responses are confidential. Although findings of this study may be published, no information that
can identify an individual will be included in any part of the study. When the data from the study is being
analyzed, the researcher, Michael Mansfield may share the survey data with his advising professors for
assistance in processing the data.

PARTICIPANT RIGHTS

Your student's participation in this study is completely voluntary. Your student can refuse to participate in the entire study or any part of the study without any negative effect on your relations with San Jose State University, Pajaro Valley Unified School District or their school. They have the right to skip any question they do not wish to answer. This consent form is not a contract. It is a written explanation of what will happen during the study if they decide to participate. They will not waive any rights if they choose not to participate, and there is no penalty for stopping their participation in the study.

QUESTIONS OR PROBLEMS

You are encouraged to ask questions at any time during this study.

- For further information about the study, please contact Michael Mansfield, 831-728-6390 ext. 6487.
- Complaints about the research may be presented to Dr. Bradley Porfilio, Director, EdD Leadership Program at San Jose State University, (408) 924-4098
- For questions about participants' rights or if you feel you have been harmed in any way by your participation in this study, please contact Dr. Pamela Stacks, Associate Vice President of the Office of Research, San Jose State University, at 408-924-2479.

SIGNATURES

Your signature indicates that you voluntarily agree to allow your child to be a part of the study, that the details of the study have been explained to you and your child, that you have been given time to read this document, and that your questions have been answered. You will receive a copy of this consent form for your records.

Participant Signature

_____

Name of Child/Minor (Printed)   Parent or Guardian Name (Printed)   Date

_____

Relationship to Child (Printed)   Parent or Guardian Signature (Printed)   Date

Researcher Statement

I certify that the minor's parent/guardian has been given adequate time to learn about the study and ask questions. It is my opinion that the parent/guardian understands his/her child's rights and the purpose, risks, benefits, and procedures of the research and has voluntarily agreed to allow his/her child to participate. I have also explained the study to the minor in language appropriate to his/her age and have received assent from the minor.

_____

Signature of Person Obtaining Informed Consent   Date

Appendix B: Student Assent Form for Survey

REQUEST FOR YOUR PARTICIPATION IN RESEARCH

THE ROLE OF SELF-EFFICACY AND GROWTH MINDSET ON MIDDLE SCHOOL MATH ACHIEVEMENT FOR AT PROMISE YOUTH: AN EXPLANATORY STUDY

Michael Mansfield, Education Leadership Doctoral Student & Brent Duckor, Ph.D. Faculty Supervisor

San Jose State University

WHAT THE RESEARCH IS ABOUT

You are invited to participate in a research study that is investigating the relationship between your beliefs about what you are capable of doing academically, your beliefs about whether or not you can grow in your academic ability, and your success in math.

PROCEDURES

If you agree to participate in the study you will be invited to participate in a brief online survey that will take place during your math class. The survey will consist of questions pertaining to your beliefs about your academic capabilities and beliefs about your ability to grow academically. These questions will be asked after you have completed the annual fall administration of the NWEA MAP assessment. The survey should take 15-25 minutes to complete. All reported information will be confidential, none of your personal information will be reported in the final study. Where data is reported, it will be reported mixed in with other students to keep everyone's individual information private. The survey will be given during the testing period, if you choose not to participate you will do the regularly assigned quiet activity directed by the teacher instead of the survey.

POTENTIAL RISKS

There are minimal risks to you that can result from this study. This study involves the NWEA MAP Assessment, which you are used to taking and survey questions about your academic abilities and belief in their ability to grow academically. These kinds of questions would be no different than typical questions a teacher may ask you during normal classroom activity. However, these survey questions, asked after taking the test, may make you feel a little uncomfortable, especially if you did not do well on the test and that made you angry or sad.

POTENTIAL BENEFITS

There are no direct benefits for students who participate in the study. The information that is provided may help researchers better understand the usefulness of the survey questions used in the survey and better understand the relationship of student beliefs about their academic abilities, student beliefs about their ability to grow academically and math achievement.

COMPENSATION

No compensation will be given for participating in this study.

CONFIDENTIALITY

Survey responses are confidential. Although findings of this study may be published, no information that can identify an individual will be included in any part of the study. We will only report information in a way that cannot be traced back to individual people. When the data from the study is being analyzed, the researcher, Michael Mansfield may share the survey data with his advising professors for assistance in processing the data. At all times your data will be protected in an encrypted, password protected computer, to protect your information.

PARTICIPANT RIGHTS

If you choose not to participate in the study you will not lose any services or privileges from the school. Your assent (agreement) is voluntary. You may refuse to participate in the entire study or in any part of the study. You have the right to not answer any questions you do not wish to answer. If you decide to participate in the study, you are free to quit at any time without any negative effect on your relations with San Jose State University, Pajaro Valley Unified School District or your school.

QUESTIONS OR PROBLEMS

You are encouraged to ask questions at any time during this study.

- For further information about the study, please contact Michael Mansfield, 831-728-6390 ext. 6487.
- Complaints about the research may be presented to Dr. Bradley Porfilio, Director, EdD Leadership Program at San Jose State University, (408) 924-4098
- For questions about participants' rights or if you feel you have been harmed in any way by your participation in this study, please contact Dr. Pamela Stacks, Associate Vice President of the Office of Research, San Jose State University, at 408-924-2479.

SIGNATURES

Your signature indicates that you voluntarily agree to participate in the study, that the details of the study have been explained to you, that you have been given time to read this document, and that your questions have been satisfactorily answered. You will receive a copy of this consent form for your records.

_____

Participant's Name (printed)

_____

Participants Signature   Date

RESEARCHER STATEMENT

In my judgment the minor/youth is voluntarily and knowingly giving assent to participate in this research study. Consent from the parent or guardian has also already been sought and obtained.

_____

Signature of Person Obtaining Informed Consent   Date

Appendix C: Construct Map—Self-Efficacy

The student believes that they can earn an A in all of their classes, including math. They believe that they can do well on all of their tests, even when they are difficult and they can master the hardest topics in all of their classes, including math. The student believes that all of the learning goals set by the teachers are achievable, including in math class. The student believes that their NWEA achievement and growth scores both demonstrate that they are capable of succeeding at math. The student believes that their growth over time chart shows them that they can succeed at math.	5	1. I am completely confident that I can earn an A in my classes. 2. I am completely confident that I can do well on all my tests, even when they're difficult. 3. I am completely confident that I can master the hardest topics in my classes. 4. I am completely confident that I can meet all the learning goals my teachers set. 5. I am completely confident that I can earn an A in my math class. 6. I am completely confident that I can do well on all my math tests, even when they're difficult. 7. I am completely confident that I can master the hardest topics in my math classes. 8. I am completely confident that I can meet all of the learning goals my math teachers set. 9. It is completely true that my NWEA growth percentile shows that I can succeed at math. 10. It is completely true that my NWEA achievement percentile shows that I can succeed at math. 11. It is completely true that my NWEA growth over time chart shows me that I can succeed at math.
The student believes that they can earn an A in most of their classes, depending on the subject and including math. They believe that they can do well on most of their tests, even when they are difficult and they can master the hardest topics in most of their classes, depending on the	4	1. I am mostly confident that I can earn an A in my classes. 2. I am mostly confident that I can do well on all my tests, even when they're difficult. 3. I am mostly confident that I can master the hardest topics in my classes. 4. I am mostly confident that I can meet

subject and including math. The student believes that most of the learning goals set by the teachers are achievable, depending on the subject and including math. The student believes that their NWEA achievement and growth scores both mostly demonstrate that they are capable of succeeding at math. The student mostly believes that their growth over time chart shows them that they can succeed at math.		all the learning goals my teachers set. 5. I am mostly confident that I can earn an A in my math class. 6. I am mostly confident that I can do well on all my math tests, even when they're difficult. 7. I am mostly confident that I can master the hardest topics in my math classes. 8. I am mostly confident that I can meet all of the learning goals my math teachers set. 9. It is mostly true that my NWEA growth percentile shows that I can succeed at math. 10. It is mostly true that my NWEA achievement percentile shows that I can succeed at math. 11. It is mostly true that my NWEA growth over time chart shows me that I can succeed at math.
The student believes that they can earn an A in some of their classes, depending on the subject and including math. They believe that they can do well on some of their tests, even when they are difficult and they can master the hardest topics in some of their classes, depending on the subject and including math. The student believes that some of the learning goals set by the teachers are achievable, depending on the subject and including math. The student believes that either their NWEA achievement or growth score demonstrates that they are capable of succeeding at math or that both only somewhat demonstrate that they are capable of succeeding at math.	3	1. I am somewhat confident that I can earn an A in my classes. 2. I am somewhat confident that I can do well on all my tests, even when they're difficult. 3. I am somewhat confident that I can master the hardest topics in my classes. 4. I am somewhat confident that I can meet all the learning goals my teachers set. 5. I am somewhat confident that I can earn an A in my math class. 6. I am somewhat confident that I can do well on all my math tests, even when they're difficult. 7. I am somewhat confident that I can master the hardest topics in my math classes. 8. I am somewhat confident that I can meet all of the learning goals my math

The student somewhat believes that their growth over time chart shows them that they can succeed at math.		teachers set. 9. It is somewhat true that my NWEA growth percentile shows that I can succeed at math. 10. It is somewhat true that my NWEA achievement percentile shows that I can succeed at math. 11. It is somewhat true that my NWEA growth over time chart shows me that I can succeed at math.
The student believes that they can earn an A in few of their classes, depending on the subject, math is not included. They believe that they can do well on a few of their tests, even when they are difficult and they can master the hardest topics in a few of their classes, depending on the subject, math is not included. The student believes that few of the learning goals set by the teachers are achievable, depending on the subject, math is not included. The student believes that either their NWEA achievement or growth score somewhat demonstrates that they are capable of succeeding at math or that both only demonstrate a little that they are capable of succeeding at math. The student believes a little that their growth over time chart shows them that they can succeed at math.	2	1. I am a little confident that I can earn an A in my classes. 2. I am a little confident that I can do well on all my tests, even when they're difficult. 3. I am a little confident that I can master the hardest topics in my classes. 4. I am a little confident that I can meet all the learning goals my teachers set. 5. I am a little confident that I can earn an A in my math class. 6. I am a little confident that I can do well on all my math tests, even when they're difficult. 7. I am a little confident that I can master the hardest topics in my math classes. 8. I am a little confident that I can meet all of the learning goals my math teachers set. 9. It is a little true that my NWEA growth percentile shows that I can succeed at math. 10. It is a little true that my NWEA achievement percentile shows that I can succeed at math. 11. It is a little true that my NWEA growth over time chart shows me that I can succeed at math.
The student does not believe that they can earn an A in all of their	1	1. I am not at all confident that I can earn an A in my classes.

classes. They do not believe that they can do well on all of their tests because they are difficult. They do not believe they can master the hardest topics in all of their classes. The student does not believe that the learning goals set by the teachers are achievable. The student believes that neither their NWEA achievement or growth score demonstrates that they are capable of succeeding at math. The student does not believe that their growth over time chart shows them that they can succeed at math.		2. I am not at all confident that I can do well on all my tests, even when they're difficult. 3. I am not at all confident that I can master the hardest topics in my classes. 4. I am not at all confident that I can meet all the learning goals my teachers set. 5. I am not at all confident that I can earn an A in my math class. 6. I am not at all confident that I can do well on all my math tests, even when they're difficult. 7. I am not at all confident that I can master the hardest topics in my math classes. 8. I am not at all confident that I can meet all of the learning goals my math teachers set. 9. It is not at all true that my NWEA growth percentile shows that I can succeed at math. 10. It is not at all true that my NWEA achievement percentile shows that I can succeed at math. 11. It is a little true that my NWEA growth over time chart shows me that I can succeed at math.

Appendix D: Construct Map—Academic Mindset

The student believes that their intelligence is something that can be developed and grown, including in math. They believe that challenges are what bring about increased ability and there is nothing that they are incapable of learning, including in math. They believe they can become good at something even if they are not naturally capable, including in math. The student believes that their NWEA achievement and growth scores both demonstrate that they are growing in their math ability. The student does completely believe their growth over time chart makes them feel like they are growing in their math ability.	5	1. It is not at all true that my intelligence is something that I can't change very much. 2. It is not at all true that challenging myself won't make me any smarter. 3. It is not at all true that there are some things I am not capable of learning. 4. It is not at all true that if I am not naturally smart in a subject, I will never do well in it. 5. It is not at all true that challenging myself in mathematics won't make me any smarter. 6. It is not at all true that there are some things in mathematics I am not capable of learning. 7. It is not at all true that if I am not naturally smart in mathematics, I will never do well at it. 8. It is completely true that my NWEA growth percentile makes me feel like I am growing in my math ability. 9. It is completely true that my NWEA achievement percentile makes me feel like I am growing in my math ability. 10. It is completely true that my growth over time chart makes me feel like I am growing in my math ability.
The student mostly believes that their intelligence is something that can be developed and grown, including in math. They mostly believe that challenges are what bring about increased ability and there is mostly nothing that they are incapable of learning, including in math. They mostly believe they can become good at something even if they are not naturally capable, including in	4	1. It is a little true that my intelligence is something that I can't change very much. 2. It is a little true that challenging myself won't make me any smarter. 3. It is a little true that there are some things I am not capable of learning. 4. It is a little true that if I am not naturally smart in a subject, I will never do well in it. 5. It is a little true that challenging myself in mathematics won't make

145

math. The student mostly believes that their NWEA achievement and growth scores both demonstrate that they are growing in their math ability. The student does mostly believe their growth over time chart makes them feel like they are growing in their math ability.		me any smarter. 6. It is a little true that there are some things in mathematics I am not capable of learning. 7. It is a little true that if I am not naturally smart in mathematics, I will never do well at it. 8. It is mostly true that my NWEA growth percentile makes me feel like I am growing in my math ability. 9. It is mostly true that my NWEA achievement percentile makes me feel like I am growing in my math ability. 10. It is mostly true that my growth over time chart makes me feel like I am growing in my math ability.
The student somewhat believes that their intelligence is something that can be developed and grown, including in math. They somewhat believe that challenges are what bring about increased ability and there are some things that they are incapable of learning, including in math. They somewhat believe they can become good at something even if they are not naturally capable, including in math. The student believes that either their NWEA achievement or growth scores demonstrates that they are growing in their math ability or that they believe that both of their scores only somewhat demonstrate that they are growing in math. The student does somewhat believe their growth over time chart makes them feel like they are growing in their math ability.	3	1. It is somewhat true that my intelligence is something that I can't change very much. 2. It is somewhat true that challenging myself won't make me any smarter. 3. It is somewhat true that there are some things I am not capable of learning. 4. It is somewhat true that if I am not naturally smart in a subject, I will never do well in it. 5. It is somewhat true that challenging myself in mathematics won't make me any smarter. 6. It is somewhat true that there are some things in mathematics I am not capable of learning. 7. It is somewhat true that if I am not naturally smart in mathematics, I will never do well at it. 8. It is somewhat true that my NWEA growth percentile makes me feel like I am growing in my math ability. 9. It is somewhat true that my NWEA achievement percentile makes me feel like I am growing in my math ability.

		10. It is somewhat true that my growth over time chart makes me feel like I am growing in my math ability.
The student believes a little that their intelligence is something that can be developed and grown, math is not included. They believe a little that challenges are what bring about increased ability and they are incapable of learning most subjects, including in math. They believe a little they can become good at something even if they are not naturally capable, including in math. The student believes a little that either their NWEA achievement or growth scores somewhat demonstrates that they are growing in their math ability or that they believe that both of their scores only demonstrate a little that they are growing in math. The student does believe only a little that their growth over time chart makes them feel like they are growing in their math ability.	2	1. It is mostly true that my intelligence is something that I can't change very much. 2. It is mostly true that challenging myself won't make me any smarter. 3. It is mostly true that there are some things I am not capable of learning. 4. It is mostly true that if I am not naturally smart in a subject, I will never do well in it. 5. It is mostly true that challenging myself in mathematics won't make me any smarter. 6. It is mostly true that there are some things in mathematics I am not capable of learning. 7. It is mostly true that if I am not naturally smart in mathematics, I will never do well at it. 8. It is a little true that my NWEA growth percentile makes me feel like I am growing in my math ability. 9. It is a little true that my NWEA achievement percentile makes me feel like I am growing in my math ability. 10. It is a little true that my growth over time chart makes me feel like I am growing in my math ability.
The student does not believe that their intelligence is something that can be developed and grown, math included. They do not believe that challenges are what bring about increased ability and they are incapable of learning most subjects, including in math. They do not believe they can become good at something even if	1	1. It is completely true that my intelligence is something that I can't change very much. 2. It is completely true that challenging myself won't make me any smarter. 3. It is completely true that there are some things I am not capable of learning. 4. It is completely true that if I am not naturally smart in a subject, I will

they are not naturally capable, including in math. The student does not believe that either their NWEA achievement or growth score demonstrates that they are growing in their math ability. The student does not believe their growth over time chart makes them feel like they are growing in their math ability.		never do well in it. 5. It is completely true that challenging myself in mathematics won't make me any smarter. 6. It is completely true that there are some things in mathematics I am not capable of learning. 7. It is completely true that if I am not naturally smart in mathematics, I will never do well at it. 8. It is not all true that my NWEA growth percentile makes me feel like I am growing in my math ability. 9. It is not at all true that my NWEA achievement percentile makes me feel like I am growing in my math ability. 10. It is not all true that my growth over time chart makes me feel like I am growing in my math ability.

Appendix E: MESH Survey

Self-Management
First, we'd like to learn more about your behavior, experiences, and attitudes related to school.

Please answer how often you did the following during the past 30 days.
During the past 30 days…

1. I came to class prepared.
2. I remembered and followed directions.
3. I got my work done right away instead of waiting until the last minute.
4. I paid attention, even when there were distractions.
5. I worked independently with focus.
6. I stayed calm even when others bothered or criticized me.
7. I allowed others to speak without interruption.
8. I was polite to adults and peers.
9. I kept my temper in check.

(Almost Never, Once in a While, Sometimes, Often, Almost All the Time)

Growth Mindset
In this section, please think about your learning in general.

Please indicate how true each of the following statements is for you:

10. My intelligence is something that I can't change very much.
11. Challenging myself won't make me any smarter.
12. There are some things I am not capable of learning.
13. If I am not naturally smart in a subject, I will never do well in it.

(Not at All True, A Little True, Somewhat True, Mostly True, Completely True)

Self-Efficacy
How confident are you about the following at school?

14. I can earn an A in my classes.
15. I can do well on all my tests, even when they're difficult.
16. I can master the hardest topics in my classes.
17. I can meet all the learning goals my teachers set.

(Not at All Confident, A Little Confident, Somewhat Confident, Mostly Confident, Completely Confident)

Social Awareness
In this section, please help us better understand your thoughts and actions when you are with other people.

Please answer how often you did the following during the past 30 days. During the past 30 days…

18. How carefully did you listen to other people's points of view?
(Not Carefully at All, Slightly Carefully, Somewhat Carefully, Quite Carefully, Extremely Carefully)

19. How much did you care about other people's feelings?
(Did Not Care at All, Cared A Little Bit, Cared Somewhat, Cared Quite A Bit, Cared A Tremendous Amount)

20. How often did you compliment others' accomplishments?
(Almost Never, Once in a while, Sometimes, Often, Almost all the time)

21. How well did you get along with students who are different from you?
(Did Not Get Along at All, Got Along A Little Bit, Got Along Somewhat, Got Along Pretty Well, Got Along Extremely Well)

22. How clearly were you able to describe your feelings?
(Not at All Clearly, Slightly Clearly, Somewhat Clearly, Quite Clearly, Extremely Clearly)

23. When others disagreed with you, how respectful were you of their views?
(Not at All Respectful, Slightly Respectful, Somewhat Respectful, Quite Respectful, Extremely Respectful)

24. To what extent were you able to stand up for yourself without putting others down?
(Not at All, A Little Bit, Somewhat, Quite A Bit, A Tremendous Amount)

25. To what extent were you able to disagree with others without starting an argument?
(Not at All, A Little Bit, Somewhat, Quite A Bit, A Tremendous Amount)

Appendix F: SEAM Survey Version 2

Construct	Item
AM#1	My intelligence is something that I can't change very much. Rev.- My smartness is something that I can't change very easily.
AM#2	Challenging myself won't make me any smarter. Rev.- Challenging myself on hard academic things won't make me any smarter.
AM#3	There are some things I am not capable of learning. Rev.- There are some academic things I am not able to learn.
AM#4	If I am not naturally smart in an academic subject, I will never do well in it. Rev.- I can only do well in academic subjects I am naturally smart in.
AM#5	My math smartness is something that I can't change very easily.
AM#6	Challenging myself in mathematics won't make me any smarter. Rev.- Working on hard things in math won't make me any smarter.
AM#7	There are some things in mathematics I am not capable of learning. Rev.- There are some things in math I am not capable of learning.
AM#8	If I am not naturally smart in Mathematics, I will never do well at it. Rev.- If I am not naturally smart in math, I will never do well at it.
AM#9	My NWEA growth percentile makes me feel like I am growing in my math ability.
AM#10	My NWEA achievement percentile makes me feel like I am growing in my math ability. Rev.- My NWEA RIT score makes me feel like I am growing in my math ability.
AM#11	The NWEA Growth Over Time Chart makes me feel like I am growing in my math ability. Rev.- The NWEA "growth over time chart" makes me feel like I am growing in my math ability.
SE#1	My NWEA growth percentile shows that I can succeed at math.
SE#2	My NWEA achievement percentile shows that I can succeed at math. Rev.- My NWEA RIT score shows that I can succeed at math.
SE#3	My NWEA Growth Over Time Chart shows me that I can succeed at math. Rev.- The NWEA "growth over time chart" shoes me that I can succeed at math.
Outcome Space—Not at All True, A Little True, Somewhat True, Mostly True, Completely True	
SE#4	I can earn an A in my classes. Rev.- I can earn A's in all of my classes.
SE#5	I can do well on all my tests, even when they're difficult. Rev.- I can do well on my tests, even when they're difficult.

SE#6	I can master the hardest topics in my classes.
SE#7	I can meet all the learning goals my teachers set. Rev.- I can meet the learning expectations my teachers set.
SE#8	I can earn an A in my math class.
SE#9	I can do well on all my math tests, even when they're difficult. Rev.- I can do well on my math tests, even when they are difficult.
SE#10	I can master the hardest topics in my math classes.
SE#11	I can meet the learning goals my math teachers set. Rev.- I can meet the expectations my math teachers set.
Outcome Space—Not at All Confident, A Little Confident, Somewhat Confident, Mostly Confident, Completely Confident	

Appendix G: SEAM Survey Version 1

Construct	Item
AM#1	My intelligence is something that I can't change very much.
AM#2	Challenging myself won't make me any smarter.
AM#3	There are some things I am not capable of learning.
AM#4	If I am not naturally smart in a subject, I will never do well in it.
AM#5	Challenging myself in mathematics won't make me any smarter.
AM#6	There are some things in mathematics I am not capable of learning.
AM#7	If I am not naturally smart in Mathematics, I will never do well at it.
AM#8	My NWEA growth percentile makes me feel like I am growing in my math ability.
AM#9	My NWEA achievement percentile makes me feel like I am growing in my math ability.
AM#10	The NWEA Growth Over Time Chart makes me feel like I am growing in my math ability.
AM#11	My NWEA growth percentile shows that I can succeed at math.
SE#1	My NWEA achievement percentile shows that I can succeed at math.
SE#2	My NWEA Growth Over Time Chart shows me that I can succeed at math.
Outcome Space—Not at All True, A Little True, Somewhat True, Mostly True, Completely True	
SE#3	I can earn an A in my classes.
SE#4	I can do well on all my tests, even when they're difficult.
SE#5	I can master the hardest topics in my classes.
SE#6	I can meet all the learning goals my teachers set.
SE#7	I can earn an A in my math class.
SE#8	I can do well on all my math tests, even when they're difficult.
SE#9	I can master the hardest topics in my math classes.
SE#10	I can meet all of the learning goals my math teachers set.
Outcome Space—Not at All Confident, A Little Confident, Somewhat Confident, Mostly Confident, Completely Confident	

Appendix H: Final SEAM Survey

Construct	Item
AM#1	My smartness is something that I can change, if I try hard in school.
AM#2	Challenging myself on difficult school work will help me learn more.
AM#3	There are many new academic things I can learn.
AM#4	I can do well in academic subjects I am naturally smart in and in those that might be difficult at the start.
AM#5	My math learning is something that I can improve with effort.
AM#6	Working on challenging math will help me learn more.
AM#7	There are many things in math I am capable of learning.
AM#8	If I am not naturally smart in math, I will never do well at it.
AM#9	My NWEA growth percentile makes me feel like I am growing in my math ability.
AM#10	My NWEA RIT score helps me understand how I am growing in my math ability.
AM#11	The NWEA "growth over time chart" helps me understand how I am improving in my understanding of math.
SE#1	My NWEA growth percentile shows that I can succeed at math.
SE#2	My NWEA RIT score shows that I can succeed at math.
SE#3	The NWEA "growth over time chart" shows me that I can succeed at math.
SE#4	I can earn A's in my classes when I try.
SE#5	I can do well on my tests, even when they're difficult.
SE#6	I can gain an understanding of difficult topics in my classes.
SE#7	I can meet the learning expectations my teachers set.
SE#8	I can earn an A in my math class.
SE#9	I can do well on my math tests, even when they are difficult.
SE#10	I can master the hardest topics in my math classes.
SE#11	I can meet the expectations my math teachers set.
Outcome Space—Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree, or No Answer	

Appendix I: Wright Maps

```
      -------------------------------------------------
                       |
   5                   |
                       |
                       |
                       |8
   4                   |
                       |
                       |
                 XXX|
   3               XX|
                   XX|
                       |
   2             XXX|
                   XX|
   1             XX|
               XXX|21
               XXX|
                XX|4 10
          XXXXXXX|9 13 14 20
   0            XX|12 16 18
                   |6 17 22
        XXXXXXXXXX|2 11 19
                XX|
          XXXXXXX|
                XX|1 15
  -1           XXX|
          XXXXXXX|5
               XXX|3 7
               XXX|
                       |
  -2                   |
                   XX|
                       |
  -3                   |
                   XX|
                       |
                   XX|
  -4                   |
                       |
                       |
  -5                   |
                       |
                       |
      =================================================
Each 'X' represents 0.6 cases
```

```
-----------------------------------------
                 |8.2
      5          |
                 |
                 |
                 |
      4          |
                 |
                 |
                 |
              XXX|
      3          |8.1
               XX|
                 |
               XX|
                 |21.2
      2          |
              XXX|4.2 10.2 20.2
                 |9.2 13.2 14.2
               XX|12.2 16.2 18.2
                 |6.2 17.2 22.2
      1        XX|2.2 11.2 19.2
              XXX|
              XXX|15.2
               XX|1.2
          XXXXXXX|
      0        XX|3.2 5.2 7.2
                 |
       XXXXXXXXXX|
               XX|21.1
          XXXXXXX|
               XX|10.1
     -1       XXX|4.1 9.1 13.1 20.1
          XXXXXXX|14.1 18.1
              XXX|6.1 12.1 16.1 17.1 22.1
              XXX|19.1
                 |2.1 11.1
     -2          |
                 |15.1
               XX|1.1
                 |
                 |3.1 5.1 7.1
     -3          |
                 |
               XX|
                 |
               XX|
     -4          |
                 |
                 |
                 |
     -5          |
                 |
                 |
=========================================
```

156

Appendix J: IRB Approval

SJSU SAN JOSÉ STATE UNIVERSITY

Office of Research
Division of
Research and Innovation

San José State University
One Washington Square
San José, CA 95192-0022

TEL: 408-924-2272
officeofresearch@sjsu.edu
sjsu.edu/research

SAN JOSE STATE UNIVERSITY
HUMAN SUBJECTS INSTITUTIONAL REVIEW BOARD

IRB Notice of Approval

Date of Approval: 8/3/2020

Study Title: The Impact of Self-Efficacy and Growth Mindset on Middle School Math Achievement for At Promise Youth: An Explanatory Study

Primary Investigator(s): Dr. Brent Duckor

Student(s): Michael Mansfield

Other SJSU Team Members:

Funding Source: None

IRB Protocol Tracking Number: 20165

Type of Review

☐ Exempt Registration: Category of approval §46.104(d)()
☒ Expedited Review: Category of approval §46.110(a)(7)
☐ Full Review
☐ Modifications
☐ Continuing Review

Special Conditions
☐ Waiver of signed consent approved
☐ Waiver of some or all elements of informed consent approved
☐ Risk determination for device:
☐ Other:

Continuing Review
☒ Is not required. Principal Investigator must file a status report with the Office of Research one year from the approval date on this notice to communicate whether the research activity is ongoing. Failure to file a status report will result in closure of the protocol and destruction of the protocol file after three years.

☐ Is required. An annual continuing review renewal application must be submitted to the Office of Research one year from the approval date on this notice. No human subjects research can occur after this date without continuing review and approval.

Approved by Dr. Pamela C. Stacks
Associate Vice President
Institutional Official
Office of Research
San Jose State University

IRB Contact
Alena Filip
Human Protections Analyst
408-924-2479
Alena.Filip@sjsu.edu

Primary Investigator Responsibilities

- Any significant changes to the research must be submitted for review and approval prior to the implementation of the changes.
- Reports of unanticipated problems, injuries, or adverse events involving risks to participants must be submitted to the IRB within seven calendar days of the primary investigator's knowledge of the event.
- If the continuing review section of this notice indicates that continuing review is required, a request for continuing review must be submitted prior to the date the provided.

Appendix K: Academic Mindset Items 1-11 (AM1-11)

Statistics	Response Categories		
	0	1	2
AM1			
Count	3	26	14
Percent	6.98	60.47	32.56
Pt-Biserial	-0.36	-0.34	0.55
Mean Location	-2.34	-0.35	0.09
Std. Dev. Of Locations	1.28	0.92	1.25
AM2			
Count	9	23	12
Percent	20.45	52.27	27.27
Pt-Biserial	-0.31	-0.28	0.6
Mean Location	-0.91	-0.48	1.2
Std. Dev. Of Locations	0.88	1.11	1.13
AM3			
Count	6	16	22
Percent	13.64	36.36	50
Pt-Biserial	-0.49	-0.22	0.55
Mean Location	-1.9	-0.43	0.61
Std. Dev. Of Locations	1.01	1.08	1.02
AM4			
Count	15	22	7
Percent	34.09	50	15.91
Pt-Biserial	-0.07	-0.21	-0.38
Mean Location	-0.33	0.52	1.38
Std. Dev. Of Locations	1.28	1.29	1.26
AM5			
Count	4	22	19
Percent	8.89	48.89	42.22
Pt-Biserial	-0.37	-0.31	0.53
Mean Location	-2.01	-0.67	0.72
Std. Dev. Of Locations	1.27	1.12	1.18

AM6

Count	13	19	12
Percent	29.55	43.18	27.27
Pt-Biserial	-0.57	0.09	0.48
Mean Location	-1.58	-0.01	1.04
Std. Dev. Of Locations	1.22	0.84	1.2

AM7

Count	8	13	23
Percent	18.18	29.55	52.27
Pt-Biserial	-0.47	-0.15	0.5
Mean Location	-1.73	-0.55	0.55
Std. Dev. Of Locations	1.49	1.11	1.11

AM8

Count	42	0	2
Percent	95.45	0	4.55
Pt-Biserial	0.1		-0.1
Mean Location	-0.18		-0.4
Std. Dev. Of Locations	1.47		1.5

AM9

Count	15	19	9
Percent	34.88	44.19	20.93
Pt-Biserial	-0.52	0.06	0.53
Mean Location	-1.13	0.03	1.31
Std. Dev. Of Locations	1.16	0.9	1.11

AM10

Count	13	25	4
Percent	30.95	59.52	9.52
Pt-Biserial	-0.39	0.07	0.5
Mean Location	-1.2	-0.12	1.88
Std. Dev. Of Locations	1.5	1.01	0.7

AM11

Count	9	21	11
Percent	21.95	51.22	26.83
Pt-Biserial	-0.23	-0.01	0.23
Mean Location	-1.18	-0.09	0.46
Std. Dev. Of Locations	1.75	1.08	1.3

Appendix L: Self-Efficacy Items 1-11 (SE1-11)

Statistics	Response Categories		
	0	1	2
SE1			
Count	11	25	8
Percent	25	56.82	18.18
Pt-Biserial	-0.35	-0.02	0.42
Mean Location	-1.12	-0.12	1.08
Std. Dev. Of Locations	1.77	1.05	0.99
SE2			
Count	14	24	6
Percent	31.82	54.55	13.64
Pt-Biserial	0.048	0.02	0.62
Mean Location	-1.27	-0.18	1.79
Std. Dev. Of Locations	1.33	0.88	0.76
SE3			
Count	13	22	7
Percent	30.95	52.38	16.67
Pt-Biserial	-0.41	-0.04	0.57
Mean Location	-1.19	-0.16	1.44
Std. Dev. Of Locations	1.45	0.84	1.21
SE4			
Count	8	20	16
Percent	18.18	45.45	36.36
Pt-Biserial	-0.28	-0.36	0.6
Mean Location	-0.76	-0.82	0.87
Std. Dev. Of Locations	1.47	0.99	1.38
SE5			
Count	12	23	9
Percent	27.27	52.27	20.45
Pt-Biserial	-0.43	-0.1	0.59
Mean Location	-1.4	-0.17	1.36
Std. Dev. Of Locations	1.43	0.83	1.32

SE6

Count	12	21	11
Percent	27.27	47.73	25
Pt-Biserial	-0.54	-0.15	0.73
Mean Location	-1.54	-0.28	1.45
Std. Dev. Of Locations	1.36	0.75	0.87

SE7

Count	11	24	8
Percent	25.58	55.81	18.6
Pt-Biserial	-0.45	-0.05	0.58
Mean Location	-1.07	-0.18	1.43
Std. Dev. Of Locations	1.05	1.14	0.98

SE8

Count	13	18	14
Percent	28.89	40	31.11
Pt-Biserial	-0.45	-0.13	0.58
Mean Location	-1.36	-0.2	0.9
Std. Dev. Of Locations	1.3	0.9	1.31

SE9

Count	15	23	7
Percent	33.33	51.11	15.56
Pt-Biserial	-0.58	0.06	0.68
Mean Location	-1.43	0.01	1.78
Std. Dev. Of Locations	1.13	0.91	0.77

SE10

Count	19	20	5
Percent	43.18	45.45	11.36
Pt-Biserial	-0.46	0.09	0.58
Mean Location	-0.87	0.11	1.91
Std. Dev. Of Locations	1.17	0.97	0.82

SE11

Count	9	24	9
Percent	21.43	57.14	21.43
Pt-Biserial	-0.45	-0.1	0.57
Mean Location	-1.14	-0.22	1.37
Std. Dev. Of Locations	1.29	1.05	0.94