False Growth Mindsets: An Exploration

Nicholas Buttrick

M.A., University of Virginia, 2016

B.A., Yale University, 2008

A Dissertation Presented to the Graduate Faculty

of the University of Virginia in Candidacy for the Degree

of Doctor of Philosophy

Department of Psychology

University of Virginia

May 2020

Committee:

Timothy D. Wilson (chair)

Sophie Trawalter

Gerald Clore

Chris Hulleman

Abstract

Incremental beliefs about intelligence (commonly known as "growth mindsets") have become a topic of intense interest among researchers, educators, and the general public. Believing that one's intelligence can grow over time as long as one tries hard and has the appropriate strategies (as opposed to believing that intelligence is a fixed quality about a person that cannot be changed, commonly known as "fixed-mindset" or entity theories) is associated with greater persistence after failure and improved academic performance. The full message isn't always fully conveyed, however, with some believing that a growth mindset simply means that anyone can succeed with hard work alone - that one should just "try, try again." This "false growth mindset" may be more prevalent than believed, and may have serious consequences. I investigate the roots and consequences of holding a false growth mindset in three studies with over 10,000 participants, including both a nationally-representative sample of 9th graders and their math teachers, and samples including nearly 4,000 nationally-representative American adults. In Study 1, I find, in a preregistered analysis, that over 38% of teachers can be characterized as holding a false growth mindset, and that students in these classrooms are more likely to view their teachers as holding an entity theory of intelligence, are more likely to hold such a belief about their own abilities, and are therefore more likely to have lower end-of year grades. In Studies 2 and 3, I then develop and validate a measure of false growth mindset beliefs, designing a scale that discriminates across a full range of potential levels, that is interpreted comparably across adults and high-schoolers, that captures a construct that is distinct from other measures of incremental thinking, and that predicts both theoretically-relevant pattern of belief and predicts societal victim-blaming above and beyond competing constructs.

To Erin Westgate, Charlie Ebersole, Samantha Heintzelman, Jordan Axt, Tim Wilson, Shige Oishi, Jerry Clore, Sophie Trawalter, Ben Converse, Brian Nosek, Jazi Brown-Iannuzzi, Stephanie Wormington, David Silverman, Calvin Lai, Liz Gilbert, Thomas Talhelm, Kelly Hoffman, David Reinhard, Katie Lancaster, Anup Gampa, Jane Tucker, Lindsay Juarez, Adi Shaked, Brandon Ng, Veronica Weser, Hyewon Choi, Meltem Yucel, Jin Bak, Diane-Jo Bart-Plange, Maura Austin, Yuching Lin, Quinn Hirschi, Remy Furrer, Andy Hales, Kosta Kushlev, Margaux Wienk, Lee Williams, Kyshia Henderson, Steph McKee, Nava Caluori, Jacob Goldstein-Greenwood, Shelly Zhang and, of course, Courtney Soderberg; all the "we's" who are hidden behind the "I's" of this dissertation.

Table of Contents

False Growth Mindsets: An Exploration

A 'growth mindset,' or the belief that one's abilities are malleable and that a person can improve them if they put in the effort, use good strategies, and are unafraid to ask for help, can lead to a range of positive outcomes throughout a life, including improving academic success and dealing with setbacks and stressors (e.g. Dweck 1999; 2006). In this dissertation, I explore a common misinterpretation of the growth mindset, known as the "false growth mindset" (Dweck, 2015; 2016a, 2016b), the belief that anyone can improve themselves through simply trying hard. While seemingly innocuous, I show that holding a false growth mindset can lead to dangerous outcomes, associated as it is with an increased belief in ideas that people are fully in control of their own outcomes and that those who have not succeeded in society are themselves to blame for their failures. Teachers holding a false growth mindset are more likely than those with a "true" growth mindset to be perceived by their students as holding the opposing belief that talents are fixed; are more likely to convince students that students' own intelligence and math ability are unchangeable; and therefore, students in these classrooms are more likely to have lower end-of-semester grades. Across two sets of studies, one (Study 1) looking at a nationally-representative sample of 9th-grade students and their math teachers, and one developing (Study 2) and content-validating (Study) a measure of the false growth mindset, I lay out some of the causes and consequences of believing that, while success is possible for all, effort is all that matters for success, and that, therefore, anyone who doesn't succeed has chosen failure.

Mindsets Matter

Growth mindsets can be quite helpful. Students who believe that their intelligence is something that can be grown over time (also known as having an incremental theory of intelligence) believe that intellectual improvement is possible if they work for it (e.g. Dweck & Leggett, 1988) and are able to reframe potential setbacks, such as academic failure experiences or challenge, as opportunities for improvement (e.g. Mueller & Dweck, 1998; though see Burgoyne, Hambrick, & Macnamara, 2020 for evidence that growth mindsets do not always correlate with theorized outcomes; and see Burnette et al., 2013 and Sisk et al., 2018 for meta-analyses with competing findings about the efficacy of growth mindset interventions).

These beliefs stand in relief against the beliefs of students with a fixed mindset (also known as an entity theory), who believe that their talents are what they are and that failure is simply an indication that they don't have what it takes to succeed. Mindset beliefs are themselves changeable, though, and interventions that teach students about growth mindsets can have real consequences for students' motivation and academic performance, even at scale (e.g. Blackwell, Trzesniewski, & Dweck, 2007; Yeager, Walton, et al., 2016; Yeager et al., 2019; Yeager et al., 2019; but see Sisk et al., 2018). By learning that one can improve with the right combination of effort, study strategies, and willingness to ask for help, students, especially those from underrepresented or minority backgrounds, seem better equipped to handle the ups-and-downs of the academic experience (e.g. Yeager & Dweck, 2012; Yeager et al., 2019).

Mindsets and Blame

Holding an incremental mindset may not be an unalloyed good, however, and may lead to blaming others for their failures, especially when those failures are repeated. This may be

grounded in fundamental differences in free-will beliefs between incremental and entity theorists. Dweck and Molden (2008), for example, review a series of studies showing that people who hold entity theories of personality tend to believe that the actions of others are more determinate – that they have less choice about their actions – and that, on the flip side, incremental theorists are more willing to believe that people have the power and ability to chart their own paths (e.g. Levy & Dweck, 1999; Levy, Stroessner, & Dweck, 1998). They argue that entity theorists tend to believe that human nature is fixed and that an individual's freedom to act is limited by their basic nature; while incremental theorists, believing in a changeable human nature and in the power of motivation, believe that people, including themselves, have a free will.

This belief in the freedom of choice has two countervailing moral implications. If someone does something morally repugnant, one can choose to blame them for it either because they freely chose the bad action or because they are inherently a bad person. Incremental theorists tend to pick the former option, while entity theorists tend to choose the latter, and indeed, when asked about the reasons for a punishment, entity theorists, believing in bad actors, are more likely to prioritize retribution, while incremental theorists, believing in bad choices, are more likely to prioritize rehabilitation (Chiu Dweck, Tong, & Fu, 1997; Gervey, Chiu, Hong, & Dweck, 1999; Yeager, Trzesniewski, Tirri, Nokelainen, & Dweck, 2011). This ascription to a fundamental nature tends to make entity theorists more vengeful, and teaching an incremental mindset may make people more generally forgiving (Yeager, Trzesniewski, & Dweck, 2013; Yeager et al., 2011).

A key part of blame is whether or not the judge believes that the perpetrator could have chosen other than they did (e.g. Alicke, 2000; Malle, Guglielmo, & Monroe, 2014; Pizarro & Tannenbaum, 2011). When bad events are made to seem less controllable, they become less blameworthy (Monroe, Dillon, & Malle, 2014). In line with classic attribution theory, which breaks down the cause of actions to factors internal to a person such as effort and ability, and factors outside of the person such as situational factors (e.g. Heider 1958), when people attribute the causes of bad outcomes to factors internal to the victim, they tend to blame those individuals (e.g. Weiner et al, 1972).

A focus on free will, however, may lead to people moralizing situations in which actors had no real choice in the first place. If one believes that all actions are free, then it may be that bad outcomes, no matter how apparently situationally determined, could have been avoided if only the victim could have chosen otherwise. In one set of studies, Americans who were primed with choice, and therefore ideas of free will, were more likely to blame people for misfortunes that were reasonably out of their control, such as having a heart attack, being physically abused, or having one's home physically collapse. Being primed with choice even led Americans to empathize less with a story about a seven-year-old boy potentially starving to death in Mali, a tragedy wholly out of his control (Savani, Stephens, & Markus, 2011). By making people think about the controllability of actions, it seems as if bad events become blameworthy ones.

Incremental theorists, with their tendency to focus on free will may be more likely than entity theorists to make this moral move. For example, in one study, those who believed that emotions are more generally controllable were more likely to have negative reactions towards a hypothetical person expressing psychological distress, as if they blamed the distressed person for

the inability to control their emotional states (Smith, 2020). Similarly, people who believe that happiness is controllable are more likely to blame the unhappy for not choosing happiness (Tullett & Plaks, 2016) and parents who believe that their children have more control over their emotions are more likely to blame them for emotional outbursts (Halberstadt et al., 2013).

Incremental theorists may also be more alert to the presence or absence of effort than entity theorists, given that a growth mindset increases both mastery orientation and an appreciation of the power of effort (Blackwell et al., 2007; Dweck, 2008, Robins & Pals 2002). In one study, for example, natural incremental theorists become more anxious when it appears that a target participant, described as a 'math geek' who is motivated to improve his verbal reasoning fails to increase his GRE verbal scores, while entity theorists are generally unfazed (Plaks, Grant, & Dweck, 2005). When it comes to their own performance, those with incremental mindsets (either dispositionally or experimentally-induced) tend to be more anxious and preoccupied by an inability to improve their own performance when given the opportunity (Plaks & Stecher, 2007). Incremental theorists, in other words, look to effort as a primary driver of outcomes, and assess its presence or absence as a key explanatory factor.

This focus on effort may cause incremental theorists to be especially judgmental in certain situations. Failure based on lack of effort is generally punished more harshly than failure based on lack of ability (Weiner, 1994), and incremental theorists, with their focus on the power of motivation, may be especially likely to use their key explanatory mechanism, effort, for individual failure. This can be seen most clearly in a set of studies that looked at how people blame others for chronic failures. If a failure happens repeatedly, it may seem as if the failing individual is, by not choosing to change, choosing to fail. Participants who were told that

personality variables such as empathy and aggression were changeable (vs. fixed) generally blamed individuals for not becoming more empathetic or less aggressive equally, as long as the failure only occurred once. But, once the failure to change became chronic, those led to believe that personality was changeable judged targets who did not change more harshly, largely because they tended to believe that the individual had the power to control their behavior and simply weren't putting in the effort to do so (Ryazanov & Christenfeld, 2018). Similarly, in a study of conflict in romantic relationships, those participants with a more incremental theory of personality were initially more optimistic about their partners' ability to change a troubling aspect of their relationship. But, as the weeks went by, and the partner failed to fully change, those participants with an incremental theory shifted, viewing their partners' failure as a voluntary choice, and therefore judging it more harshly (Kammrath & Peetz, 2012). By focusing on the lack of effort the target put in, incremental theorists become stricter and less forgiving.

This relationship between incrementalism and blame for repeated failure is not unequivocal. Researchers, for example, have shown that, while those who believe that one's weight is something that can be controlled are more likely to blame the obese (including themeselves) for their inability to lose weight, they are also more charitable towards those with weight issues, because they believe that they still might be able to lose weight in the future (Burnette et al., 2017; Hoyt et al., 2017). This double-edged-sword of incrementalist thinking can be honed through interventions that address the issue of blame while still maintaining the efficacy beliefs inherent to a growth mindset. A message that stresses the use of appropriate strategies and downplays self or other-blame, can lead to reduced anti-fat prejudice and

improved body image (Hoyt et al., 2019; see also Burnette et al., 2019 for similar findings in the domain of addiction).

The False Growth Mindset

These concerns about effort may be magnified by a potentially-common misunderstanding of the growth mindset. In the canonical form of the growth mindset, students should be learning that people can improve their academic performance as long as they try hard, use the proper strategies, and are unafraid to ask for help (Dweck, 2008). The traditional approach to teaching growth mindsets (e.g. Yeager et al., 2012), however, stresses the power of hard work for improvement above and beyond other factors. With its stress on "growing the brain," it is entirely possible that students may come away from the intervention believing that effort alone is the key to improvement, and that, simply, if you try hard, you'll do well. This so-called "false growth mindset" (Dweck, 2015; 2016a, 2016b) may not carry the same motivational benefits as the "true growth mindset," and may even hinder one's performance.

As evidence that the false growth mindset can be prevalent and pernicious, Wormington et al. (under review) analyzed the responses to a prompt at the end of a large-scale growth mindset intervention in a community-college developmental math class. Students assigned to the growth-mindset induction were asked to write a short passage about what they had learned from the intervention so that other students, who had not been taught about the growth mindset, could potentially improve their own performance. This exercise should elicit what students believe to be the key elements of what they had learned. These essays were hand-coded for whether they mentioned nine key concepts: that intelligence is malleable; that people can learn or succeed; the importance of effort; that, with work, someone can achieve success; that practice is important;

that challenge leads to opportunities for growth; that failure is an opportunity to change; that one needs a good strategy; and that asking for help is ok.

People were then assigned to one of three profiles based on their combination of responses. Around a fifth of the respondents mentioned no elements of the growth mindset and generally appeared to be disengaged with the material as a whole. About a third of the sample wrote about most or all of the elements of the growth mindset, doing a good job of describing the full panoply of the growth mindset. The remaining half of the sample, however, is probably best categorized as succumbing to the false growth mindset. These people believed that intelligence was malleable and that one could get better with effort, but failed to mention anything at all about the importance of good strategies or help-seeking. Importantly, what participants took away from the intervention had consequences: compared to the control condition, which received no growth-mindset training, only those who wrote about the full growth mindset improved their grades by the end of the semester. Those who appeared to learn nothing or only appeared to understand the false growth mindset saw no benefits of the mindset intervention.

Those students who believed that simply hard work was the only strategy worth following may have had little to fall back upon when faced with repeated setbacks. This may be seen as a sort of metacognitive failure, or an inability to effectively track whether one's strategies are working, and if they are not, switching tasks to try a different approach (e.g. Schneider and Pressley, 1989). If people know only one way to approach a problem, it can be hard to even understand that they need to think of alternatives because the idea of approaching a problem differently may simply never occur.

Tweaking a mindset intervention to focus more time on the importance of strategy may make the intervention more effective. Compared to the traditional intervention, a revised intervention that deemphasizes effort in favor of strategy and help seeking increases challenge seeking, decreases performance avoidance, and appears to influence end-of-year grades to a similar degree (Yeager, Romero, et al., 2016).

It is possible that the false growth mindset may also arise independently of an intervention, as a part of a worldview that endorses meritocratic beliefs – beliefs that hard work automatically leads to success (e.g. Ledgerwood, Mandisodza, Jost, & Pohl, 2011), or as part of the Protestant Work Ethic, which identifies hard work as the chief virtue in life (e.g. Furnham, 1984), fused with the individualistic belief that one's outcomes in life are purely under one's control. Beliefs that the world is just, and that people get what they deserve (e.g. Lerner, 1980; reviewed in Hafer & Bègue, 2005) have, as a component, a belief that those who put in the effort are able to reap the reward – a belief functionally similar to the false growth mindset itself. Those who believe in a meritocratic society or in the Protestant Work Ethic are more willing to blame lower-status people for their lower status in the world, as opposed to the systems which may have had a role to play (e.g. Cozzarelli, Wilkinson, & Tagler, 2001; Crandall, 1994; Katz & Hass, 1988), and may even be willing to blame themselves for their own bad outcomes (Hafer & Olson, 1993; McCoy & Major, 2007). The false growth mindset, working, as it does, through similar channels, may have similar global effects on empathy for the struggles of others.

The Mindsets of Others

Mindsets aren't just an intrapsychic phenomenon – the expectations that others have for us affects our own outcomes. Most profoundly, of course, this can be seen in the literature on

self-fulfilling prophecies (e.g. Rosenthal & Jacobson, 1968; reviewed in Madon, Willard, Guyll, & Scherr, 2011; but see also Jussim & Harber, 2005). Perceiver expectations for target outcomes can shape targets' own expectations, as when teacher expectations for success leads students to believe that they are able to succeed. Beliefs in how one can succeed may be similarly transmissible – the mindsets of parents, for example, powerfully shapes the mindsets of their children (Haimovitz & Dweck, 2016), the generally accepted beliefs of an organization can shape the mindsets of those who wish to become a member (Murphy & Dweck, 2010), and teachers may be able to intentionally instill growth mindsets in their students (Ferguson, Phillips, Rowley, Friedlander, 2015).

People can accurately infer the mindsets held by important people in their life, such as teachers or significant others, and an inference that one's evaluator has a fixed mindset can lead to self-doubt in academic settings (Reich & Arkin, 2006). This set of inferences can seriously impact student performance – in a large-scale study of STEM professors, for example, classes taught by professors with fixed mindsets had twice as large of a racial achievement gap as classes taught by professors with growth mindsets (Canning, Muenks, Green, & Murphy, 2019). On the one hand, it may be that teachers with more of a fixed mindset shape their classroom in line with these beliefs, which affects student performance directly. One longitudinal study, for example, demonstrated that teachers who more strongly held entity theories of intelligence were more likely to emphasize performance-oriented educational practices (such as doing well on tests, regardless of strategies), which led to an increase in student entity theorizing throughout the semester (Park, Gunderson, Tsukayama, Levine, & Beilock, 2016; see also Leroy, Bressoux, Sarrazin, & Trouilloud, 2007).

Teachers with different mindsets may also treat students differently. There is evidence, for example, that the implicit biases of teachers can leak out into the expectations they have for their students, with concomitant effects on student grades (van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010). Teacher attributions for success and failure shapes the emotions they feel towards the failing student and the behaviors they take in turn. If a teacher interprets a student's failure as coming from a lack of ability, they may be driven to pity, while if they interpret the failure as coming from a lack of effort, they may be driven to anger, and the less a teacher feels that their student's outcomes are something that the teacher can control, the less help they're willing to give (Butler, 1994; Georgiou, Christou, Stavarinides, & Panaoura, 2002). Teachers may provide more support and instruction to students who they believe have a high potential for success, while proving fewer resources for those who they feel have less potential (Fuchs, Fuchs, & Phillips, 1994). Teachers with a fixed mindset, therefore, may focus more on students who they think have the talent to succeed in their class, writing off those who they feel cannot improve.

It can be demotivating as a student to think that your teacher doesn't believe in you. Teacher practice can change student's attitudes as deeply as students' content knowledge (reviewed in Blazar & Kraft, 2017), and, for example, students taught by professors who they believe to have a fixed mindset feel higher degrees of impostor-syndrome and increased worries that they might not belong in their STEM classes. These beliefs can lead to less engagement with their classes, including lower rates of attending class, higher rates of class dropout, and lower end-of-year grades (Muenks, Canning, Green, Zirkel, Garcia, & Murphy, under review; see similar findings in Rattan, Savani, Komarraju, Morrison, Boggs, & Ambady, 2018).

False growth mindsets may be especially problematic in educational settings. Students whose teachers mainly praise their effort, without praising their strategies have been shown to interpret that praise as demeaning, indicating that their teacher doesn't believe that they have the ability to succeed in the class and that the teacher is just praising their effort as a sort of 'consolation prize.' Students who make this attribution are then more likely to internalize the message that they're hearing, leading them to question their abilities and ironically shading them towards an entity theory of their own intelligence (see Amemiya & Wang, 2018 for a review). And if a teacher believes that all outcomes are under individual control, then failing students, regardless of the causes of their failure may be interpreted as having failed due to problems within themselves, and thus may be greeted with anger and disavowal. These teachers may, therefore, be unusually intolerant of failure, especially of repeated failure, and this may lead, via the way they set up their classrooms and the messages which their teaching conveys, to students internalizing the sense that failure says something about their abilities, which is the classic hallmark of the fixed mindset.

The Current Research

In this paper, I investigate the prevalence, consequences, and correlates of the false growth mindset, using a variety of measurement techniques and a large and diverse sample of Americans to understand how a false growth mindset might work. Are false growth mindsets real? Are they a problem? And how do you measure them?

In Study 1, I assess the belief structure of a sample of 9th-grade mathematics teachers, using patterns of self-reports to categorize them as possessing true growth mindsets, false growth

mindsets, or entity theories of intelligence. I next look at the effects that these teacher beliefs have on the students in their classrooms, demonstrating that having a teacher with false growth mindset leads to students increasing their belief that their teacher has a fixed mindset and decreasing their own growth mindset beliefs, both of which predict lower end-of-year grades.

In Studies 2 and 3, I design and validate an instrument to directly measure a false growth mindset, assessing beliefs about the importance for success of both effort and appropriate strategy use. The newly-developed scale has a repeatable two-factor structure, is interpreted similarly across both groups one would expect (across age and gender) and across groups where invariance is not a given (adults, college students, and high-schoolers), and has moderate test-retest stability. The measure correlates well with other related constructs, is distinct from other measures of growth mindset or incremental theories of personhood, predicts theoretically-important patterns of responding to other scales, and can predict societal victim-blaming above and beyond competing measures.

Study 1: The Real-World Impact of the False Growth Mindset

While anecdotal evidence suggests that misunderstandings of growth mindsets may be more widely-held than education researchers are comfortable with (e.g. Sorhagen, personal communication), I know of no systematic effort to quantify their prevalence or consequences. Little is known about how teachers understand growth mindsets as a multifaceted construct; about how students' perceive and interpret their teachers' false growth mindsets; about how closely teachers' and students' perceptions of teachers' false growth mindsets align; about the motivational effects of teachers' false growth mindsets on their own teaching practice; about the

motivational effects of teachers' false growth mindsets on their students' beliefs; and about how the interrelationship between teacher false growth mindset and students' own mindsets affect students' performance in the classroom.

In the present study, I fill in a number of these gaps. I use teachers' self-reported practices and beliefs to understand how they think about the growth mindset, constructing a set of mindset profiles to assess how teachers naturally understand the various aspects of the growth mindset. I then look at their students, measuring student beliefs about their teacher's mindset as well as students' own mindsets, to understand how teacher mindsets get transmitted into student beliefs. Finally, I investigate how teacher mindsets influence student performance in the classroom, both directly and mediated through changes in the ways that students come to understand their teachers and themselves.

My analyses are unusually generalizable. The data come from a preregistered analysis of data collected in the National Study of Learning Mindsets (NSLM), a large nationally-representative sample of 9th grade students and their mathematics teachers (containing responses from over 12,000 students in over 350 classrooms across over 70 American public schools), allowing me to suggest that these patterns of belief may be widely shared across American teachers and schools.

I predicted that teacher mindset would matter, and that having a teacher with a false growth mindset (compared to a true growth mindset) would lead, among their students, to an increase in entity theorizing, increased perceptions that the teacher cares about effort, and an increase, ironically, in perceptions that their teacher has a fixed mindset about ability. These increases, I predicted, would be borne out in students' end-of-year GPA, such that these student

beliefs would mediate the relationship between teacher mindset and end-of-year GPA. I additionally predicted that these relations would be stronger for students that came into the year in a more precarious place, grade-wise, as students with lower GPAs in previous semesters would be the most likely to struggle, and therefore the most likely to be victim-blamed by teachers with a false growth mindset.

Method

Disclosures

Preregistration. All models were refined on an initial 10% of the data, randomly selected, that was made available as an exploratory set for this purpose. After models were finalized, the remaining 90% of the data were made available for the confirmatory tests reported below. The preregistration for this project can be found at https://osf.io/vfxds?view_only=87d1b260a5994cd4bc863b2ef2680d36

Data & Materials. All data and codebooks for the National Study of Learning Mindsets are available at https://doi.org/10.3886/ICPSR37353.v1. Analysis scripts for this project can be found at https://osf.io/qz4g9/?view_only=87d1b260a5994cd4bc863b2ef2680d36.

Participants

Data come from the National Study of Learning Mindsets (NSLM). The NSLM is a large nationally-representative study of a growth-mindset intervention conducted with over 12,000 9th graders from 65 public schools across the United States along with over 350 of their mathematics teachers. The NSLM is the largest randomized-controlled-trial of growth-mindset interventions to date in a US K-12 setting, and its careful sampling provides an unmatched window into

processes surrounding growth mindsets (see Yeager et al., 2019 for more about the NSLM, and see Gopalan & Tipton, 2018 and Tipton, Yeager, Iachan, & Schneider, 2019 for more about the stratified random sampling and national representation). At the beginning of the semester, students completed a short online session, in which they either learned about a growth mindset, or received a control intervention. 1-4 weeks later, all students completed a set of follow-up measures. Teachers, blind to the condition that their students were randomized into, were simply surveyed at the beginning of the semester. For further details about the NSLM, see Yeager et al., 2019.

The subsample of data analyzed here come from the pairing of students and teachers where the student filled out the full complement of measures below, where the student and teacher could be uniquely matched (i.e. the student only had one teacher who filled out survey measures in the dataset), and where end-of-year grades for the student were provided by the school. Depending on the precise specification, the data comprise between 3,835-5,453 students nested within 305 teachers in 61 schools.

Materials

Teacher Survey. To assess teachers' beliefs about the importance of effort, ability, strategies, and their attitude towards more and less successful students, I selected the following questions from the beginning-of-the-semester Teacher Survey:

Teachers were given a set of videos of students being taught in classrooms, and were asked "Based on your professional judgment, what percent of the math students in the last three videos probably had the intellectual potential to excel at the highest levels of high school math, like Calculus?"

They were then told to "Imagine that one of your 9th grade math students was very discouraged in math class. The student kept getting low grades on assignments. The student didn't always try, but when he or she did try hard, the student would still get things wrong, even after practicing." They were given a space to freely respond with the sorts of feedback that they would provide, and were asked (on a scale from 1 = extremely likely to 5 = not at all likely) how likely they would be to say the following statements: "Don't worry—it's okay to not be a math person;" "Please come get tutoring after class/school;" "Keep working hard and you'll get it;" "Let's look at what went wrong in your process and see what happens when we fix it;" and "Let's see what you don't understand and I'll explain it differently."

Next, teachers were told to "Imagine one of your math students was doing very well in math class. The student is getting really high grades on assignments, often without trying or putting in much time. The student doesn't ask questions because he or she isn't confused by very much." They were given a space to freely respond with the sorts of feedback that they would provide, and were asked (on a scale from 1 = extremely likely to 5 = not at all likely) how likely they would be to say the following statements: "Let's find something to challenge and confuse you, so you can learn more;" "When it's easy, that's when it's time to try something harder;" "Great job, you must be working hard;" "It's great that it's so easy for you;" and "You're lucky that you're a math person."

Teachers were then asked about their general practices in 9th grade math (on a scale from 1 = extremely true to 5 = not at all true): "I tell my 9th grade students it is important to work hard in math class;" "I try to put my slower/remedial 9th grade students together for group work;" "I

allow my 9th grade students to revise and resubmit work when they did not get a good enough score initially;" and "It slows my class down to encourage lower achievers to ask questions."

Finally, teachers were asked about their beliefs about ability more generally and their attitudes about teaching more specifically (on a scale from 1 = strongly disagree to 6 = strongly agree, with no neutral midpoint): "People have a certain amount of intelligence, and they really can't do much to change it;" "Being a top math student requires a special talent that just can't be taught;" "If you want to succeed in math, hard work alone just won't cut it; you need to have a natural gift or talent;" "Some people are just born great teachers; if you're not, there's not much you can do to become a really great teacher;" and "If I really try hard, I can get even the most difficult or unmotivated student to learn."

We additionally made use of the free-responses to both the struggling and excelling students, using coding generated by Alex Browman and his team (Browman, Miele, O'Dwer, & May, under review). The authors had two independent coders rate the teachers across three dimensions (on a three point bipolar scale with a midpoint indicating neither of the two options): the degree to which the teacher was autonomy-supportive versus controlling (e.g. "Acknowledges or prompts a dialogue/discussion about their rationale or the unique way in which they have chosen to think about or approach their work" versus "Proposes to make the student do the work in the teacher's own/preferred way"; ICC for struggling student response = .78; ICC for excelling student response = .72); the degree to which the teacher was mastery-oriented versus performance-oriented (e.g., "Emphasizes the importance of understanding course material" versus "Emphasizes the importance of getting the right answer or of not making mistakes on course work"; ICC for struggling student response = .77; ICC for

excelling student response = .83); and the degree to which the teacher expressed positivity versus negativity (e.g., "Expresses warmth, approval, encouragement, or gave positive feedback to the student" versus "Expresses frustration, annoyance, or hostility or gave negative feedback to the student"; ICC for struggling student response = .80; ICC for excelling student response = .86). See Browman et al. (under review) for more details.

Student Survey. I used three sets of variables from the student surveys, which I each collapsed into a composite variable when used in analyses. Composite variables were based on the average of non-missing data for the scale in question. All values, unless otherwise noted, come from surveys taken roughly 1-4 weeks after the teacher surveys. In the pre-pregistration exploratory dataset, I conducted an exploratory factor analysis, using parallel analysis to determine the number of factors to extract, and found that my items measuring student perceptions about their teacher's beliefs formed two factors. In the confirmatory dataset, I found the same pattern of results (TLI = .964, RMSEA = .044, 43% of variance explained). See Table 1 for factor loadings and see the Supplement for the correlation matrix between the student-level variables.

Student Growth Mindset. This construct (see e.g. Yeager, Romero, et al., 2016) was generated from three questions (all on a scale from 1 = strongly disagree to 6 = strongly agree with no neutral midpoint): "You have a certain amount of intelligence, and you really can't do much to change it;" "Your intelligence is something about you that you can't change very much;" and "Being a 'math person' or not is something that you really can't change. Some people are good at math and other people aren't." Scale alpha = .80 [.79, .80], $M = 2.70$, $SD = 1.20$.

Table 1
Loadings for Student Perceptions of Teacher Beliefs, Study 1

Item	Beliefs about Ability	Beliefs about Effort
My math teacher thinks that some kids are smart and others are not.	**.74**	.08
My math teacher seems to like you better if you are good at math.	**.69**	.04
My math teacher seems to believe that only a few students will understand the hardest problems.	**.67**	.02
My math teacher calls you smart if you are good at math.	**.52**	-.24
My math teacher seems to believe students can't really change how smart they are.	**.49**	-.03
My math teacher asks questions to be sure we are following along when s/he is teaching.	.05	**.68**
My math teacher accepts nothing less than our full effort.	-.08	**.62**
My math teacher believes that everybody in my class can be very good at math.	.15	**.60**
My math teacher thinks failure helps us learn and grow.	-.07	**.48**

Student Perceptions of Teacher Beliefs About Ability. This construct was generated from five questions (all on a scale from 1= extremely true to 5 = not at all true): "My math teacher seems to believe that only a few students will understand the hardest problems;" "My math teacher seems to like you better if you are good at math;" "My math teacher calls you "smart" if you are good at math;" "My math teacher seems to believe that students can't really

change how smart they are;" and "My math teacher thinks that some kids are smart and others are not." Scale alpha = .74 [.73, .75], *M* = 3.91, *SD* = 1.01.

Student Perceptions of Teacher Beliefs About Effort. This construct was generated from four questions (all on a scale from 1 = not at all true to 5 = extremely true): "My math teacher believes that everybody in my class can be very good at math;" "My math teacher thinks failure helps us learn and grow;" "My math teacher accepts nothing less than our full effort;" and "My math teacher asks questions to be sure we are following along when s/he is teaching." Scale alpha = .69 [.68, .70], *M* = 3.69, *SD* = 0.84.

Student Grades. My primary DV was end-of-year student grade point average (standardized to a 0-4.3 scale). For those students who took the intervention in the fall semester, I used an average of their fall and spring semester GPAs, while for those students who took the intervention in the spring semester, I just used their spring-semester GPA. *M* = 2.45, *SD* = 1.24.

Analysis Strategy

Analyses were conducted in two phases. In phase one, I identified patterns of response across teachers, using these patterns to profile them as having a true growth mindset, or an entity theory of intelligence. Once teachers had been classified to a profile, I then looked at the outcomes of students in their classrooms, analyzing whether teacher mindset affected student grades and student beliefs.

To categorize teachers' beliefs, I used a multi-level non-parametric latent profile analysis (LPA), with teachers nested within schools (Finch & French, 2014; Henry & Muthen, 2010). A latent profile analysis is a person-centered analytic approach which looks at the way that participants respond to a set of items, modelling the natural variation in patterns of responding as

the function of a set of distinct underlying latent variables. This approach synthesizes across a set of variables, identifying differences in how they relate to each other across people and then looks at how patterns of responding within people predicts outcomes of interest. This contrasts with a more variable-centered approach, which would look at the relationship of each variable singly or, at best, in terms of interaction-terms, and which would have trouble conceptualizing the relationship of more than two or three variables jointly (after which one is left interpreting 4- and 5-way interactions; see Bergman & Trost, 2006).

For example, researchers can analyze students' self reports of various aspects of internal and external motivation and instead of looking at the effects, say, of just one measure of internal motivation in predicting academic performance, perhaps moderated by a second measure (as in a traditional variable-centered approach), they can instead use a person-centered approach to build profiles of responding (such as those students high on all kinds of motivation, or those who have relatively low levels of internalized guilt-induced motivation but high levels of goal-induced external motivation) and then use those profiles themselves to identify what real-world (not just statistically-artifactual) patterns of motivation are related to the most positive outcomes (Wormington, Corpus, & Anderson, 2012). These latent profiles allow researchers to capture entire complex worldviews, which makes it a good match for identifying a construct such as the false growth mindset which is defined in terms of the relative relationships of its constituent parts. Additionally, by adding a multilevel aspect to these analyses, I can model how these underlying latent variables express themselves differently across higher-level units, in this case allowing us to look at how the way that profiles are expressed across teachers differ based on the schools in which they teach.

As LPA is an inherently exploratory framework, where models are chosen that best fit the data, I fit multiple potential models and registered a decision rule about which model to interpret in further analyses. To select the number of profiles to model, I first fit a series of flat LPAs (ones with no nesting) that varied in their number of profiles. Based on the fit of the flat models, I then fit multilevel non-parametric LPAs that varied in their number of profiles both at Level 1 (the teacher level) and at Level 2 (the school level). To identify the best-fitting model, I selected the model with the lowest sample-adjusted BIC (Bayesian Information Criterion), as long as each profile contained at least 10% of the sample, so as to make sure that each profile was capturing a meaningful proportion of the overall data (see Gaspard et al., 2019 for a broadly similar decision rule). If the model with the lowest aBIC did not generate profiles that each contained at least 10% of the sample, I selected the model with the next highest aBIC, etc.

After identifying the best-fitting model, I then interpreted the profiles, classifying them as true growth mindset theorists, false growth mindset theorists, or entity theorists, based on their patterns of results and how they related to my theoretical framework, looking, within each profile, at item raw scores and the interrelationship between the items. For the preregistered analyses, I created a dummy variable to allow me to compare teachers with a false growth mindset (coded as 1) against teachers with a true growth mindset (coded as 0).

Once teachers had been categorized into profiles of responding, I collapsed students across their intervention condition, first determining whether students in classes where their teacher had a false growth mindset ended up with worse end-of-semester grades than those students in classes where their teachers had a true growth mindset, by fitting a multilevel regression predicting students' final grades from the teacher mindset dummy described above,

with a random intercept for math classroom. In this, as in the other models presented here, there are no additional covariates above what is described in the text.

I then followed-up by testing whether the impact of teacher false growth mindset was stronger on students with lower grades in the previous semester by fitting a multilevel regression predicting students' final grades from the teacher mindset dummy interacted with students' prior-semester grades, with a random intercept for math classroom.

Finally, to look at whether student beliefs mediated the link between teacher mindset and student grades, I fit a 2-(1, 1, 1)-1 multilevel structural equation model with random slopes (Stride, Gardner, Catley, & Thomas, 2015; Preacher, Zhang, & Zyphur, 2011; Preacher, Zyphur, & Zhang, 2010). I modeled a level-2 (classroom) manifest independent variable (the teacher mindset dummy) predicting three level-1 (student-level) mediating manifest variables: students' time-2 beliefs about their teacher's growth mindset; students' time-2 beliefs about their teacher's theory of effort; and student's own time-2 mindset. Both the IV and the mediators then predicted a manifest level-1 DV: students' end of semester grades. For all the paths from the level-1 mediators, I fit both random slopes and random intercepts. Code for the model can be found at https://osf.io/8tvc5/?view_only=87d1b260a5994cd4bc863b2ef2680d36. I additionally tried to fit a matching 2-(1, 1, 1)-1 model with moderation by prior-semester grades, but the model would not converge. All LPA and SEM models were run in MPlus 8, while all data cleaning and regression models were run using R 3.6.0. All multilevel models were conducted using the lme4 and lmerTest packages, with p-values calculated using Satterthwaite-approximated degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2017).

Results

Confirmatory Analyses

Profiling Teacher Beliefs. In this analysis, I sought to identify which, if any, teachers possessed patterns of belief that matched my picture of the false growth mindset. I had predicted that there would be a set of teachers who strongly believed that people could become smarter and better at math, but who would focus mainly on the importance of effort, downplaying flexible strategy use or asking for help, and who would therefore create differing classroom structures built around these beliefs. Using the full complement of teacher measures and data described above from 305 teachers within 61 schools, I initially fit flat (un-nested) models that ranged from 2 profiles all the way to 8 profiles. Based on the fit statistics of the flat models, I then fit a set of multilevel nonparametric models with 2 to 5 Level-1 (teacher) profiles and 1 to 5 Level-2 (school) profiles. Based on the aBIC and coverage of the profiles, I settled on a solution with three Level-1 profiles and one Level-2 profile (aBIC = 19,666.39, AIC = 19,608.21, Entropy = .89). The existence of only one Level-2 profile indicated that the three lower-level profiles did not differ based on the school that the teacher taught in, and that latent profiles were expressed similarly regardless of school context. See Table 2 for profile fit solutions, and see Figure 1 for a graphical representation of the final profile solution.

Table 2
Fit statistics for non-parametric multi-level latent profile analyses, Study 1

# of Level 1 Profiles	# of Level 2 Profiles	Parameters	LL	AIC	BIC	aBIC	Entropy	Smallest Profile Proportion
2	1	79	-9823.17	19804.33	20098.24	19847.69	0.96	0.38
2	2	81	-9823.17	19808.33	20109.68	19852.79	0.82	0
2	3	83	-9823.17	19812.33	20121.12	19857.88	0.59	0
2	4	85	-9823.17	19816.33	20132.56	19862.98	0.55	0
2	5	87	-9823.17	19820.33	20144.00	19868.08	0.68	0
3	1	106	-9698.11	19608.21	20002.57	19666.39	0.89	0.22
3	2	109	-9698.11	19614.21	20019.73	19674.03	0.55	0
3	3	112	-9698.11	19620.21	20036.89	19681.68	0.47	0
3	4	115	-9698.11	19626.21	20054.05	19689.32	0.41	0
3	5	118	-9698.11	19632.21	20071.21	19696.97	0.38	0
4	1	133	-9603.78	19473.57	19968.37	19546.56	0.92	0.079
4	2	137	-9603.78	19481.57	19991.25	19556.75	0.66	0
4	3	141	-9603.78	19489.57	20014.13	19566.95	0.92	0
4	4	145	-9603.78	19497.57	20037.01	19577.14	0.48	0
4	5	149	-9603.78	19505.57	20059.89	19587.34	0.83	0
5	1	160	-9523.52	19367.03	19962.28	19454.84	0.93	0.033
5	2	165	-9549.07	19428.15	20042.00	19518.70	0.73	0
5	3	170	-9523.52	19387.03	20019.49	19480.33	0.70	0
5	4	175	-9549.07	19448.15	20099.20	19544.19	0.68	0
5	5	180	-9549.07	19458.15	20127.80	19556.93	0.86	0

Note: LL = Log-likelihood, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, aBIC = Adjusted Bayesian Information Criterion

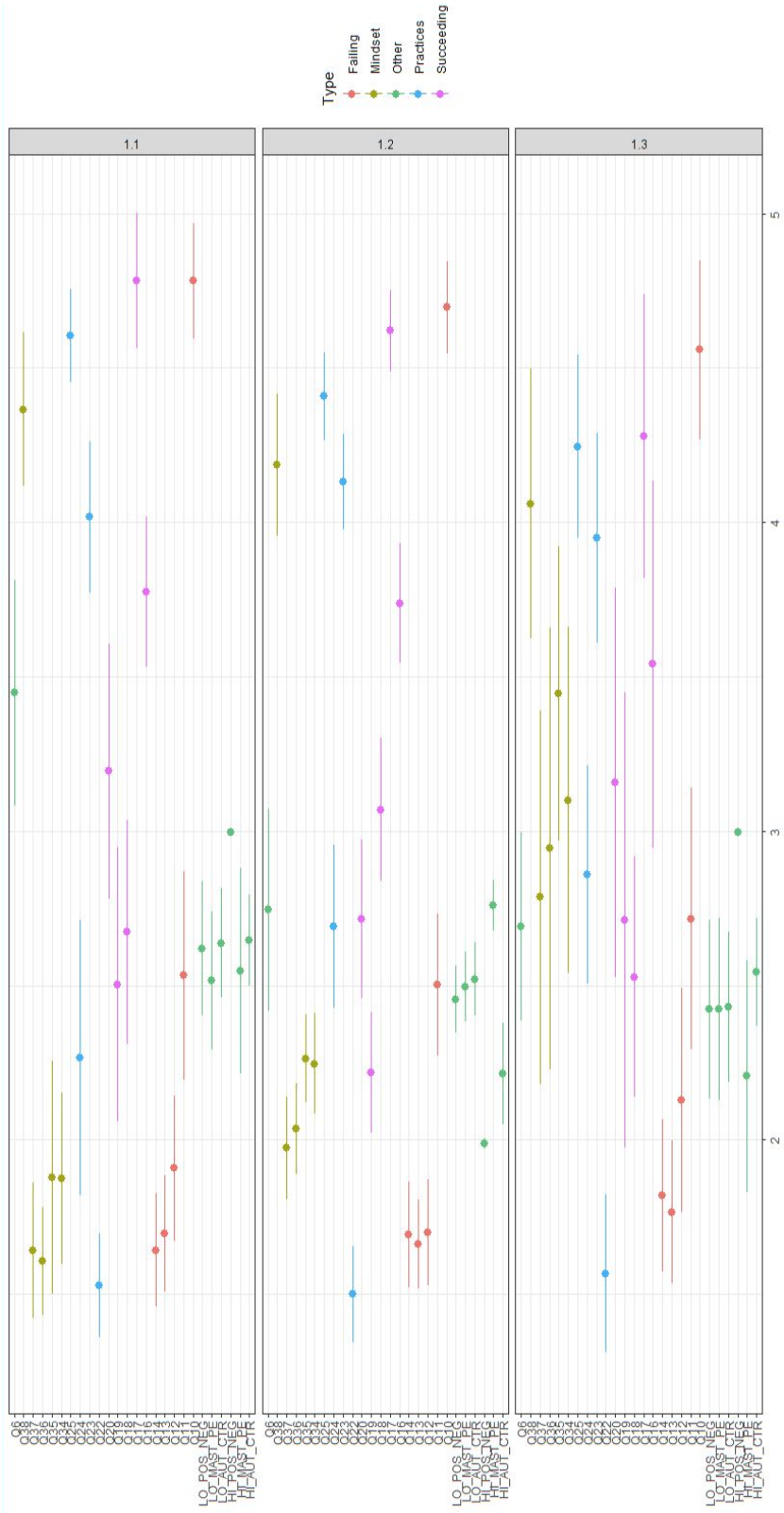


Figure 1. Final profiles from non-parametric multi-level latent profile analysis. Colors indicate types of questions asked of the teachers.

Teachers belonging to Profile 1, which I characterized as exhibiting a False Growth Mindset (117 teachers, or 38% of the sample), *strongly* agreed with the statement that people could grow their ability, and that any student had the intellectual potential to do well at the highest level of college mathematics. This group was more likely to praise the efforts of successful students while, at the same time, being less likely to push them to try harder challenges. Coding of the teachers' free responses about what they would say to the struggling and successful students suggested that these teachers were also more likely to respond in more authoritarian fashion to struggling students: demanding that they do things the way that the teacher wanted and being less likely to acknowledge the student's way of seeing the world; while tending towards strong positivity in their messages to the succeeding students. In other words, these teachers appeared to believe that students can grow their ability and that anyone had the potential to succeed, but they focused more on students' effort, and less on helping students find strategies that work for them.

Teachers belonging to Profile 2, which I characterized as exhibiting a True Growth Mindset (120 teachers, or 39% of the sample), were more measured in their growth mindset beliefs, agreeing that people could grow their abilities, but not unreservedly. These teachers were less likely to praise the effort of succeeding students, and provided feedback that was more empathetic and more supportive of each student's individual needs and worldviews. These teachers, in other words, planned on providing behavioral support for success, with an eye towards being mindful of the different problems that different students may be having.

Teachers belonging to Profile 3, which I characterized as exhibiting an Entity Theory of Intelligence (68 teachers, or 22% of the sample), tended to believe that one's intelligence is

fixed, that being a top math student is the sort of thing that cannot be taught, and that success in math requires talent, not just hard work. These teachers were also more likely to believe that teaching itself is something that requires talent, and that really great teachers are born, not made. In short, these teachers reported beliefs that ability is fixed for students, for themselves, and for their fellow teachers. These teachers also provided a wide range of free responses to both the struggling and succeeding students, underlining the heterogeneity that can underly the fixed mindset.[1]

Surprisingly, teachers did not differ across profiles in their self-reported classroom practices. Teachers from all three profiles appeared to be equally likely to allow students to resubmit work, or to agree that it slowed down their class to let lower achievers ask questions. I return to this point in my Discussion.

Assessing the Direct Relationship Between Teacher Mindset and Student Grades. After settling on a profile solution I used profile membership to predict the grades of students in those classrooms. I had predicted that students whose teacher held a false growth mindset would have lower end-of-semester grades than those whose teacher held a true growth mindset, especially if the student struggled in the prior semester. Using data from 5453 students nested within 139 teachers (those student and teacher pairs for which I were able to uniquely match classrooms, and for whom I had student grade data available), I did not find evidence that students with teachers holding false growth mindsets had lower end-of-year grades than those

[1]I additionally ran sets of models with a more restricted number of self-report items, using just questions referring to hypothetical students, and to just questions referring to hypothetical failing students, but I found that the best-fitting models for those specifications were harder to theoretically interpret. See the Online Supplement for those fit statistics and profile solutions.

students with teachers holding true growth mindsets, False M = 2.46 [2.31, 2.62]; True M = 2.56 [2.41, 2.72], b = -.10 [-.31, .12], t(128.55) = -0.91, p = .37, d = -0.079 [-0.13, 0.24].

I additionally found no evidence that the effect of teacher mindset on student grades was different based on students' prior-semester grades, interaction b = 0.0002 [-0.06, .06], t(1399) = 0.009, p = .99. This null effect is somewhat hard to interpret however, as it is based on a far smaller sample, just 1403 students nested within only 30 teachers. Nearly ¾ of the overall sample did not have prior-semester grades reported, and only 10 of the profiled teachers were categorized as believing in a false growth mindset. Those students that that did have prior-semester grades reported, moreover, had significantly higher end-of-semester grades than those who did not: t(3249.1) = 5.45, p < .001, d = 0.15 [0.09, 0.20], potentially suggesting that schools that were able to produce prior-semester grades may be qualitatively different that schools that were not so able, further limiting the comparison.

Assessing the Mediational Role of Student Beliefs in the Relationship Between Teacher Mindsets and Student Grades. Next, I looked at the relations between a teacher's mindset and a student's belief about their own ability and their beliefs about what the teacher thinks about ability, and whether those psychological constructs themselves predict end-of-year student grades. I had predicted that student beliefs and perceptions would mediate the link between teacher mindset and student outcomes, with students whose teacher held a false growth mindset more likely to think that their teacher had an entity theory of intelligence, that they cared especially strongly about students' efforts, and that students themselves would come to believe that they had a fixed ability to do well in the class, all of which would lead to lower student grades. Using data from 4905 students nested within 138 teachers (additionally restricting the

data in the direct tests to those students for whom I could calculate self-report beliefs), I found that a teacher holding a false growth mindset predicted an increase in time-2 student entity-theory beliefs (*a* path), $b = 0.21$ [0.056, 0.36], $p = .007$; and an increase in student entity-theory beliefs predicted a decrease in end-of-year grades (*b* path), combined Level-1 & mean Level-2 $b = -1.73$ [-2.94, -0.52], $p = .005$; overall indirect effect = -0.36 [-0.70, -0.014], $p = .041$. While I did find that students with teachers who had a false growth mindset were less likely to think that their teacher had a growth mindset about ability, $b = -0.12$ [-0.24, -.002], $p = .046$, I did not find evidence for mediation of teacher false growth mindset to end-of-year student grades through student beliefs about their teacher's theories of ability, indirect effect = 0.18 [-0.14, 0.51], $p = .27$; nor through student beliefs about their teacher's theories of effort, indirect effect = 0.70 [-0.24, 0.10], $p = .41$. See Figure 2 for a simplified path diagram.

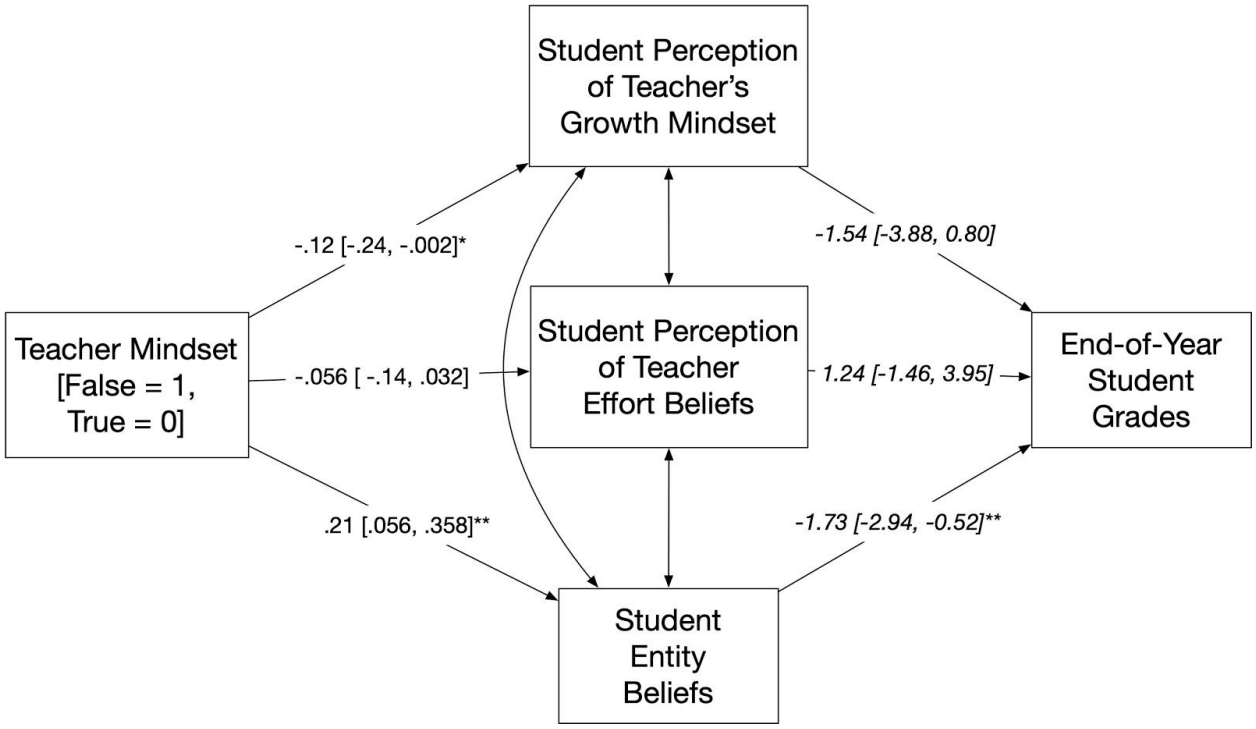


Figure 2. Simplified output for 2-(1,1,1)-1 Structural Equation Model.

Italics indicate combined Level-1 and Level-2 effects. Brackets contain 95% confidence intervals; *p <. 05, **p < .01, ***p< .001.

Exploratory Analyses

Alternate Tests of Mediation. To further check the results of the multilevel SEM, I ran an additional set of mediational models outside of an SEM framework, testing whether changes in student entity theories mediated the effect of teacher false growth mindset on end-of-year student grades without taking student beliefs about their teacher's theories of effort and intelligence into account. Using the *mediation* package in R (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014), I fit a mediational model with random intercepts for classroom (excluding the random slopes because the package cannot yet handle models of that type) on data from 4,914 students nested within 138 teachers (the sample differs from the SEM models due to missingness in the beliefs about teacher effort and ability theories), and found evidence that student entity theories did still mediate the relationship: teacher false growth mindset predicted student entity theory (*a* path): b = 0.19 [0.043, .34], t(123.48) = 2.53, p = .013; and, controlling for teacher mindset, student entity theory predicted end-of-year grades (*b* path): b = -0.14 [-0.17, -0.12], t(4875.41) = -10.65, p < .001; average causal mediation effect = -0.026 [-0.050, -0.010], p = .01.

I additionally tested whether student perceptions of teacher growth mindset mediated the relationship between teacher mindset and student grades, without controlling for students' own mindset beliefs. Using data from 4,620 students nested within 138 teachers, I found that teacher false growth mindset predicted lowered student perception of teacher growth mindset beliefs (*a* path): b = -0.13 [-0.24, -0.013], t(126.29) = -2.18, p = .031; and, controlling for teacher mindset, student perceptions of teacher growth mindset predicted end-of-year grades (*b* path): b = 0.11

[0.076, 0.14], $t(4561.49) = 6.70$, $p < .001$; average causal mediation effect = -0.013 [-0.027, 0.00], $p = .026$.

Finally, as an attempt to identify the causal direction of the effect of student mindsets - whether teachers with a false growth mindset changed the mindsets of their students, or whether classrooms full of students with entity theories led to teachers adopting false growth mindsets of their own, I fit a model predicting student's time-2 entity theories from their time-1 entity theories (measured at the beginning of the semester, 1-4 weeks before the time-2 survey) and the teacher mindset dummy. Using data from 4,880 students nested within 138 classrooms, controlling for students' time-1 mindsets and with a random intercept for classroom, teacher false growth mindset marginally predicted students' time-2 mindsets (*a* path): $b = 0.075$ [-0.010, .16], $t(117.70) = 1.73$, $p = .087$. Controlling for time-1 student and teacher mindsets, however, time-2 student mindsets still predicted end-of-year grades (*b* path): $b = -0.13$ [-0.16, -0.093], $t(4795.85) = -7.71$, $p < .001$. The effect of time-1 student mindsets on end-of-year grades, controlling for teacher mindset and time-2 student mindsets was far weaker: $b = -0.035$ [-0.069, -0.0011], $t(4798.46) = -2.02$, $p = .044$). The overall indirect effect for the mediation of teacher mindset predicting end-of-year grades through time-2 student mindset, controlling for time-1 student mindset, was marginally significant: average causal mediation effect = -0.0094 [-0.021, 0.00], $p = .062$. In other words, it appears that students' mid-semester mindsets are the most predictive of end-of-year grades, and that those mindsets changed marginally more with teachers who held a false growth mindset, suggesting that it is more likely that teachers are affecting student mindset beliefs than the reverse.

Alternate Grade Specifications. I then re-ran these models looking just at students' end-of-year math GPA (calculated in the same way as my end-of-year measure above). My results were largely consistent. As with total end-of-year GPA, I found no direct effect of teacher mindset on student grades (5542 students nested within 140 teachers): b = -0.091 [-.31, .13], t(128.02) = -0.82, p = .41; and no moderation by prior GPA (1423 students nested within 32 teachers): b = -0.0024 [-0.059, 0.063], t(1418.08) = 0.077, p = .94.

In my SEM models (4592 students nested within 139 teachers), I again found evidence for a significant indirect effect for teacher mindset predicting student grades through student entity theorizing: -0.32 [-0.64, -0.054], p = .048, while finding no significant evidence for an indirect effect through either measure of student perceptions of their teachers.

Finally, in a causal mediation framework, I again found evidence for the mediating effects of student entity theorizing on the relationship between teacher mindset and end-of-year math grades (5002 students nested within 139 teachers): average causal mediation effect = -0.025 [-0.048, 0.00], p = .028; and again found evidence for mediation through student perception of teacher mindset beliefs (4703 students nested within 139 teachers): average causal mediation effect = -0.014 [-0.027, 0.00], p = .012.

Full details are reported in the Online Supplement, as are two additional sets of analyses that use alternate end-of-year grade specifications: one set of models that uses just total spring GPA for all students, and another set that used total spring GPA where it was available, imputing in the mixed spring/fall total GPA from the main analysis where it was not. My point estimates and p-values change somewhat, mainly as a function of fluctuating sample size (the two alternate grade measures correlate with the original measure at r(5689) = .968 [.965, .969], p < .001 for

just the spring GPA; and $r(6781) = .973$ [.972, .975], $p < .001$ for the spring/imputed GPA), but the conclusions I take away largely do not.

Comparison of True Growth Mindset Theorists with Entity Theorists. As a robustness-check for the profile analysis, I ran an additional set of models comparing teachers classified as true growth mindset theorists against teachers classified as entity theorists, mirroring the more typical analysis of the effect of teacher beliefs on the beliefs and outcomes of their students (e.g. Canning et al., 2019). As above I found no main effect of teacher mindset on end-of-year student grades, $b = 0.069$ [-0.22, 0.36], $t(96.21) = 0.47$, $p = .64$, using 3950 students nested within 101 teachers. I did, however, find results consistent with theory (using 3605 students nested within 103 teachers), where students whose teachers held a true growth mindset were themselves directionally less likely to endorse entity theories at time-2 (*a* path): $b = -0.19$ [-0.39, 0.0081], $t(82.30) = -1.88$, $p = .064$. These changes in student mindset were meaningful, as those students who had stronger entity theories (controlling for teacher mindset) had lower end-of-year GPA (*b* path): $b = -0.15$ [-0.18, -0.12], $t(3577.13) = -9.26$, $p < .001$; with an overall indirect effect of teacher true growth mindset predicting end-of-year GPA mediated through student mindset that was marginally significant: average causal mediation effect = 0.028 [-0.00078, 0.06], $p = .056$. See Online Supplement for the SEM analyses, which show similar results.

Comparing Student Growth Mindsets Across Teacher Profiles. Finally, I compared student entity theories at Time 2 across the three profiles. Using 6,133 students nested within 170 teachers, I ran a one-way ANOVA with a random intercept for classroom, and found that students nested within the three teacher mindset profiles differed significantly: $F(2, 141.79) =$

3.48, $p = .034$. Follow-up uncorrected pairwise tests (Maxwell & Delaney, 2004) showed that students in classrooms where their teacher had a true growth mindset ($M = 2.58$, $SD = 1.18$) endorsed entity beliefs significantly less than students in classrooms where their teacher had a false growth mindset ($M = 2.78$, $SD = 1.22$): $b = -0.19$ [-0.38, -0.0081], $z = -2.44$, $p = .015$; and students where their teacher had a true growth mindset endorsed entity beliefs marginally less than students in classrooms where their teacher had an entity theory ($M = 2.73$, $SD = 1.16$): $b = -0.19$ [-0.42, 0.046], $z = -1.88$, $p = .060$. There was no difference between students in classrooms where their teachers held false growth mindsets versus those where their teachers held entity theories: $b = 0.0046$ [-0.23, 0.24], $z = 0.046$, $p = .96$. See Figure 3.

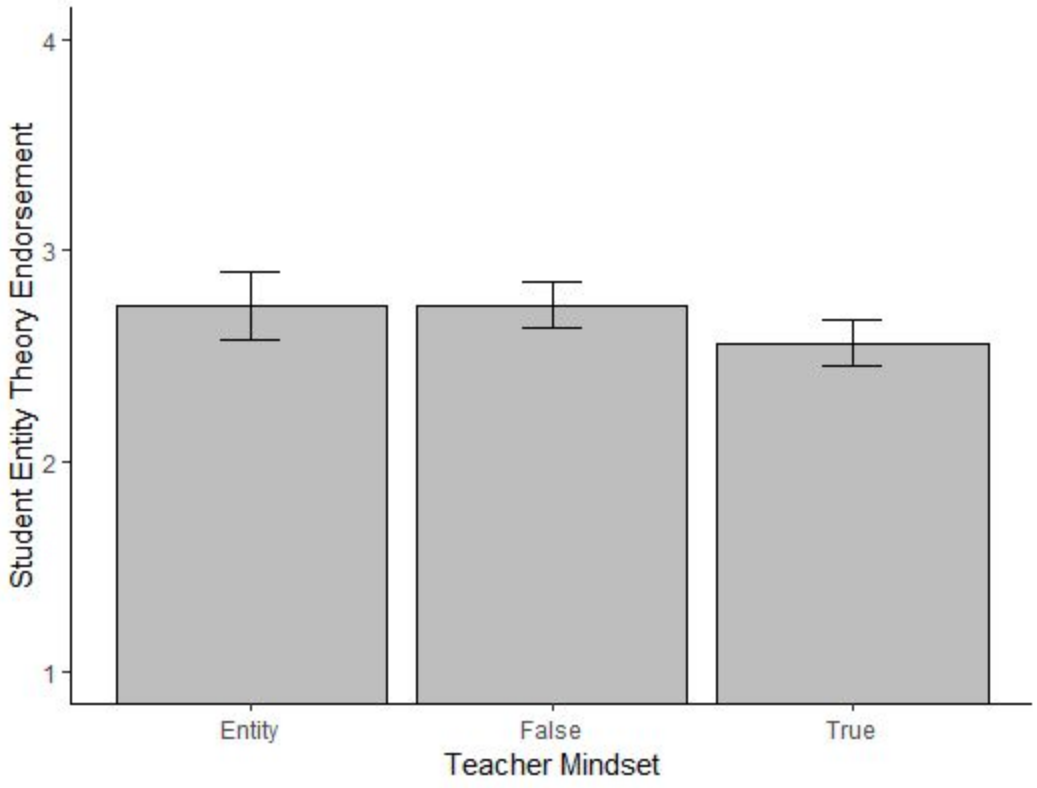


Figure 3. Student endorsement of entity theories of intelligence by the mindset profile of their teacher. Error bars indicate 95% confidence intervals.

A similar analysis looking at differences in student perceptions of teacher mindsets found that students whose teachers had a false growth mindset perceived their teachers to have less of a growth mindset ($M = 3.83$, $SD = 1.04$) than those students with teachers holding a true growth mindset ($M = 3.96$, $SD = 1.00$) or an entity theory ($M = 3.96$, $SD = 0.98$), albeit only marginally significantly, $F(2, 143.98) = 2.39$, $p = .095$.

Discussion

Growth mindset interventions teach that people can improve their abilities through hard work, the use of good strategies, and a willingness to ask for help. A misunderstanding of the message, however, may omit the last two elements, leaving holders of such a "false growth mindset" with a sort of bumper-sticker version - that anyone can improve anything, as long as they simply try (and that, therefore, if someone isn't getting better, they're simply not trying hard enough). I predicted that a) teachers with these beliefs would hold harsher views of students that failed to succeed in their classrooms, which would b) be perceived by their students as a belief that their teacher held a fixed mindset and would c) cause students in those classrooms to think that their own abilities were fixed, and that these would together d) lead to lowered end-of-year grades, especially amongst those students who came into the year with lower prior-year GPAs, who would be the most likely to struggle.

In a set of preregistered analyses, using data from a large nationally-representative sample of public-school 9th grade mathematics classes, I conducted a multilevel latent profile analysis of a rich array of teacher self-reports and behavior, and found that there are a substantial

proportion of teachers who believe that everyone should be able to succeed but who focused on effort and effort-related feedback without emphasizing the importance of challenge or flexible strategy-use. 38% of teachers surveyed were categorized as holding this pattern of belief - what I identify as the false growth mindset - and another 22% of teachers held a pattern of belief that matched an entity theory of intelligence, agreeing with the idea that being a top math student or a good teacher is a talent, the sort of thing that cannot be taught, the sort of thing a person can be born to do, and that some people can do these things and that others simply cannot. Only 39% of teachers appeared to hold a pattern of belief that could be considered a true growth mindset, with an awareness that people can improve their academics, but that it is important to provide supportive, tailored feedback, and that each student may have different sorts of needs. These patterns were consistent across all the schools I had data for (we only fit one Level-2 profile), suggesting that the false growth mindset is expressed in similar ways across teachers, regardless of the educational setting that they find themselves in.

Students in these classrooms had very different sorts of experiences based on the belief-profile of their teachers. Compared to students with teachers holding true growth mindsets, those students in a classroom with a teacher holding a false growth mindset were more likely, a month or so into the semester, to identify that teacher as having entity theories of ability, and were more likely to hold an entity theory of their own abilities, while beliefs about their teachers' beliefs about the importance of effort did not differ. These beliefs about ability predicted lower end-of-year grades, mediating the link between teacher beliefs and student outcomes. In exploratory follow-up analyses, I found that the endorsement of entity theories among students

with teachers holding false growth mindsets was essentially equivalent to those among students with teachers holding entity theories.

Contrary to expectations, however, I did not find evidence for direct effects of teacher mindset on student grades or that teachers with a false growth mindset were any harsher on students more likely to struggle in their class, as prior-semester grades did not moderate the relationship between teacher mindset and end-of-year grades. While I can only speculate about the lack of direct effect, I suspect that had I created submodels for those students with underrepresented identities I may have found different patterns than for those with more dominant identities, as prior work has shown, both experimentally (Yeager et al., 2020) and meta-analytically (Sisk et al., 2018) that growth mindset interventions work most strongly on those who are underrepresented; and that teacher mindsets have the strongest effects on underrepresented students (Canning et al., 2019). Due to analytic issues with fitting multigroup multilevel SEM models, however, I have not yet tested this possibility empirically in my data. I would, however, interpret the result finding no moderation by prior-semester grades with some caution, as due to an unexpectedly-large amount of missing data, compounded by a bias in the missingness, I may not have had the appropriate sample or power to properly test this hypothesis.

I also expected that my growth mindset profiles would differ in their reported practices. However, these did not appear to be delineating variables, with all teachers endorsing similar approaches in their teaching, and there was relatively little difference across my profiles in how teachers reported that they would address a student struggling in their classroom. In both cases, I may be dealing with an issue of "cheap talk," where teachers find it easier to think about their teaching-self in the abstract, without having to deal with the concrete everyday, where they may

not be able to live up to their ideals (e.g. Thompson, 1984). It is costless to tell researchers about the more-intensive teaching practices one would like to do but it may be far harder to put them into practice, with all the real-world tradeoffs that inevitably ensue. It may be instructive, then, that I appeared to see greater differentiation in the coding of the free-responses to the students, where the teacher had to generate responses themselves, without experimenter-cued answers to fall back upon.

The distinction between talk and action highlights an area for further research. Recent studies suggest that the transmission of mindsets from authority figures to learners may be more complicated than originally theorized, and that the mindsets of authorities don't always predict the mindsets of learners (e.g. Park et al., 2016; see Haimowitz & Dweck, 2017 for a review). It may be that what a teacher personally believes and the actions that they take in the classroom don't always line up; that they fail to connect abstract beliefs with motivational practices. In one qualitative study, for example, it was shown that four teachers, who all equally endorsed growth-mindset beliefs, had practices that often diverged from their self-reported beliefs and that their students, therefore, were receiving very mixed messages (Sun, 2019). This discrepancy between what one professes and what one does may be explained through a more holistic approach to understanding growth mindset beliefs, and that those with a false growth mindset, focusing on effort, may be setting up classrooms differently from those with a true growth mindset, focusing more on strategy use. While I found no such differences in this study, a properly observational approach, looking at what teachers *do*, rather than what they *say*, may prove useful.

Reinforcing the importance of studies of complex belief, not just basic self-report, the vast majority of teachers in this study would certainly self-identify as having a growth mindset, but as the analyses show, precisely how that mindset is comprised and instantiated has very different repercussions for students. This issue marks the importance for mindset educators to focus on intervention fidelity, making sure that interventions are delivered appropriately and as-intended (e.g. Hulleman & Cordray, 2009; Murrah, Kosovich, & Hulleman, 2017; O'Donnell, 2008; and see Burnette, in prep, for a newly-developed checklist for growth mindset interventions, specifically). Mis-specified interventions may bias people towards a misunderstanding of the meaning of the growth mindset, and researchers have shown that what one takes away from a growth mindset intervention alters its effects: students who interpret a growth mindset as involving effort alone showing no improvement in end-of-year grades relative to a control while those who interpret a growth mindset as involving flexible strategy use and a willingness to ask for help do show improvement (Wormington, under review). Future work investigating how different growth mindset interventions lead to differing levels of both true and false growth mindsets is a vital step in ensuring that the benefits of the true growth mindset are properly unlocked. After all, those teachers classified as holding a false growth mindset have students who look very similar to those teachers endorsing an entity theory, and look very different from those holding a true growth mindset.

In closing, I note one primary, unavoidable issue with these analyses: that I was unable to directly measure a false growth mindset, relying instead on a proxy measure. The profiles that I identified do differ from each other in meaningful ways, but as with factor analysis, for example, it is we, the researchers, who are post-hoc interpreting the psychological roots of these

differences. While I identified and labeled the profiles prior to the regression and SEM phase of my analyses, and while my profiles do largely act in a predictable way (with teachers classified as possessing a true growth mindset inculcating a stronger growth mindset in their students than those teachers classified as entity theorists, for example), they are nevertheless dependent on the set of questions selected for analysis and may not represent true sets of beliefs in the broader population. Looking more deeply at false-mindset beliefs, with a psychometrically-validated tool designed for the purpose, therefore, is a clear need, and one that I turn to next.

Study 2: Scale Structure

For a scale to provide meaningful information, it should be grounded in theory, it should differentiate across the full continuum of theorized responses, it should reliably and repeatedly index the same relationship among its component parts, it should have no known biases across demographic groups, and, if the scale is designed to measure a stable trait, it should be able to replicate its performance in the same set of people over time, all while being relatively efficient to deliver (e.g. Clark & Watson, 1995; Messnick, 1995; Tay & Jebb, 2018). To design a scale, I first gathered related constructs and generated a set of potential items. I then reduced the number of items using iterated exploratory factor analyses, followed by Item Response Theory analysis. In a separate sample, I confirmed that the factor structure was stable and that it was largely invariant across two demographic groups - that the scale behaved in the same way across men and women, and across those older and younger than my median age. I then investigated whether the scale was invariant across groups who might be expected to see the world differently:

high-schoolers, college students, and adults. Finally, in a third sample, I assessed its stability over time, using a two month test-retest.

Method

Participants

I recruited three separate samples for these analyses, one for testing and developing item structure, one for determining the limits of my findings, and one for assessing test-retest reliability. In the first sample, I (as part of a broader project) worked with a survey company to recruit a non-probability sample of adult Americans, ostensibly recruited so that overall sample resembled the adult population of the United States (according to the most recently available Current Population Survey, conducted by the U.S. Census Bureau) on the following demographics: gender, age, education, ethnicity (Hispanic vs. not), race (allowing each respondent to select more than one race), Census region, and income. 1598 participants agreed to have their data analyzed (M age = 51.66, SD = 11.75; 64.8% female), which I broke into two separate subsamples based on when they completed the survey, with 751 participants in the first (exploratory) subsample and 847 participants in the second (confirmatory) subsample.

In the second sample, I used data collected by the Character Lab Research Network, a service that allows researchers to run studies in middle and high-school classrooms. The scale was delivered to 358 high school students, of which 312 passed manipulation/data checks (52.2% 9th graders, M age = 14.82 years, SD = 1.31 years, 54.1% female).

In the third sample, I used data from the Pretest of a psychology department participant pool at a large public university, in which the scale had been embedded (n = 966, although only

598 participants completed the scale, for a completion rate of 60.0%), and then followed up two months later with a subsample of the original sample (n = 220, *M* age = 18.63, *SD* = 0.89; gender = 74.6% female). Due to missing responses, however, only 122 participants completed the scale at both timepoints.

Materials

To develop the scale, I began by collecting related constructs and generating lists of possible items, both within the authorship team and by reaching out to topic experts. See Table 3 for the initial items generated; the full list of related topics can be found at https://tinyurl.com/vop4qc4. I gave participants in the exploratory subsample all 31 items in a random order, measured on a scale from 1 = strongly disagree to 7 = strongly agree.

Table 3
Initial Items for False Growth Mindset Scale, Study 2

In order to overcome a challenge, all you have to do is try your best.
If people just work hard, they can get what they want.
People's outcomes are determined by their own actions
There is a clear link between hard work and success
If someone does not achieve their goal, it is because they did not try hard enough.
If someone does well on a challenging task, it must be because they worked really hard on it.
Successful people are the people who work the hardest.
Anyone who is willing to work hard enough is able to find a decent job.
I can always obtain outcomes that are important to me by working hard.
If people work hard enough, they can be whatever they want to be in life.
The harder you work at something, the better you will be at it.
If you don’t work hard and put in a lot of effort, you probably won’t do well.
If people work hard they almost always get what they want.
The best way to complete a challenging task is by trying it over and over.
If at first you don't succeed, try, try again
I am a stubborn person
Setbacks don’t discourage me.
If someone does well on a challenging task, it must be because they found a way to do it that works well for them.
If someone does well on a challenging task, it must be because they found the best way to do it.
The best way to complete a challenge must differ from one challenge to another
Successful people are the people who work the smartest.
I know what to do if my initial plan does not work well

I keep track of my process and, if necessary, I change my techniques or strategies
When someone is stuck on something, it is important to find a new approach
I think less of a person when they ask for help
If I cannot complete a task, I will ask someone for help.
I would rather deal with problems by myself
I don't like to ask others for help unless I have to
I like to get advice from others before I make a decision
When you aren't able to complete a task, it is important to ask someone for help
I'd feel better about myself knowing I didn't need help from others.

Analysis Strategy

I began by conducting a set of iterated exploratory factor analyses (EFA) on the items (using oblimin rotations) to form factors, dropping items that loaded singly onto factors or that had strong cross-loadings, then looked at the latent space coverage of the remaining items, using Item Response Theory (IRT), dropping additional items that provided relatively limited unique information. I then used the confirmatory sample to fit a confirmatory factor analysis (CFA), testing whether the proposed model acceptably fit the data. Following well-established conventions, I used a CFI > .95 and an RMSEA < .06 to indicate a well-fitting model (Hu & Bentler, 1999).

After assessing overall fit, I conducted three sets of measurement invariance tests to determine whether the scale was operating in the same way in my various samples. In the first two tracks, I used the confirmatory part of my adult sample, looking at whether the scale performed differently across men and women, and across the median age of my sample (with those younger than 56 binned as ‘young’ and those 56 or older binned as ‘old). These sets of comparisons, across groups that are expected not to differ, act as a validity check for the scale, ensuring that there is no hidden measurement invalidity (see Hussey & Hughes, 2020 for more on this approach). In the third track, I compared across groups of theoretical interest, investigating invariance across the adults of the confirmatory sample (collapsed across age and gender), college students (collected from the Departmental Pretest) and high-schoolers (collected from the Character Lab sample).

Measurement invariance tests use multigroup structural equation modelling (SEM) to fit a set of increasingly constrained models, assessing whether enforcing increasing levels of

similarity between the models leads to meaningful changes in model fit (e.g. Brown, 2006; Millsap, 2011; Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). The first step in a measurement invariance approach is to test whether members of the different groups cluster the same set of items with the same latent constructs; that is, whether the basic factor structure fits in all groups (configural invariance). To test this model, I fit separate submodels for each group, using the same factor structure as the base CFA, and tested whether this multi-group model indicated misfit.

The second step is to test whether the indicators load on to their latent factors equivalently across groups - whether each item contributes to its factor in the same way for each group (metric invariance). To test this model, I constrained factor loadings and the covariance between the two latent factors to be equal across groups, and tested for misfit relative to the configural models.

The third step is to test whether mean-level differences in the indicators are represented equivalently in the latent means - whether the scale means can be interpreted in the same way across groups (scalar invariance). To test this model, I additionally constrained the intercepts of each indicator to be equivalent across groups, testing for misfit relative to the metric invariance models.

Finally, the fourth step is to test for full invariance - whether the residuals of the models are different across the groups - whether the scale captures equivalent variance in both groups. To test this model, I additionally constrained the residuals of the models, testing for misfit relative to the scalar invariance models.

If models indicated relative misfit at any level, I tested for *partial invariance*, looking to see which items are causing the misfit, then freeing that element to differ between the groups and continuing on in the steps of invariance testing with the other elements of the scale (see, e.g. Bryne et al., 1989).

Because the results of nested ΔX^2 tests are sensitive to pragmatically-meaningless differences between models as sample sizes grow larger, and because my sample is relatively large, I instead used a change in alternate fit indices to determine comparative model misfit (Cheung & Resnvold, 2002). Following recommendations, I use a $\Delta CFI < .02$ to determine misfit of scalar-invariance models, and a $\Delta CFI < .01$ to determine misfit of all other models (Cheung & Resnvold, 2002; Rutkowski & Svetina, 2014; though see Meade, Johnson, & Braddy, 2008, who propose a cutoff of $\Delta CFI < .002$; and see Little, 2013 for a critique of this stricter cutoff level).

Finally, I assessed test-retest agreement, whether scores at time-1 have similar values at time-2 (e.g. Polit, 2014; Revelle & Condon, 2019), across two months by calculating the intraclass correlation between time-1 and time-2 scores on the scale between raters. In selecting across the many varieties of ICC, I chose ICC(2, *k*), as I had multiple raters and were interested in generalizing to a broader population (Shrout & Fleiss, 1979). To assess the relative degree of reliability, I used cutoffs of ICC > .9 indicating excellent reliability; ICC > .75 indicating good reliability; ICC > .5 indicating moderate reliability; and ICC < .5 indicating poor reliability (Koo & Li, 2016; Portney & Watkins, 2000).

Results

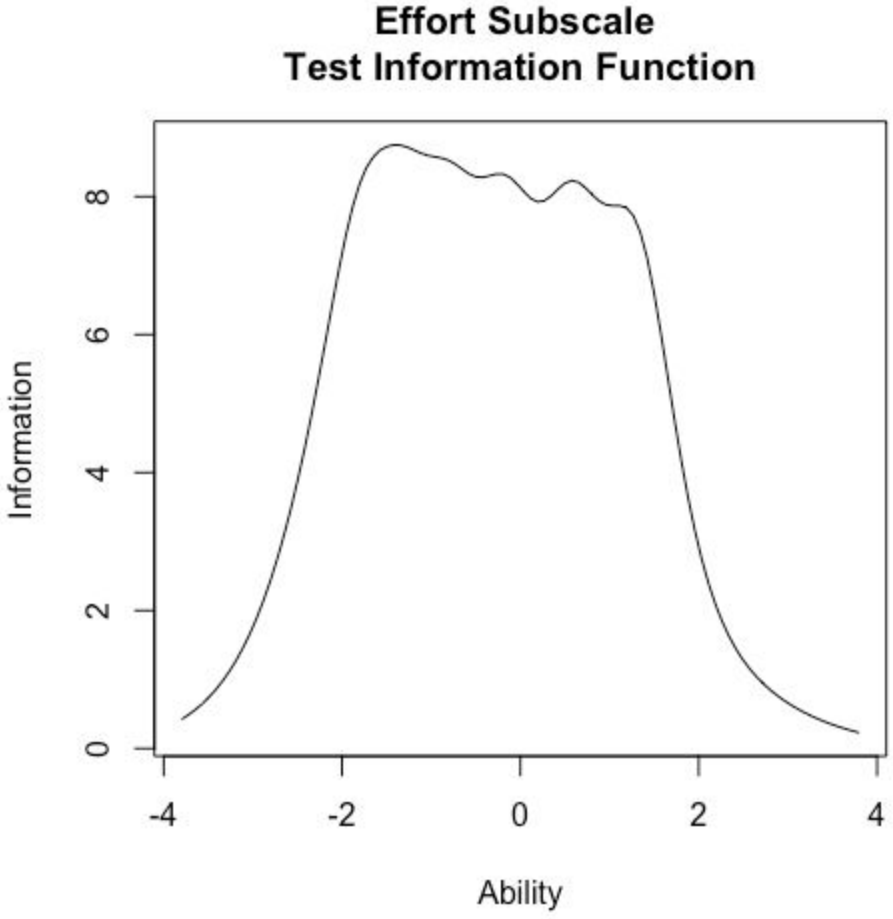
Exploratory Factor Analysis

In the initial 31 items, parallel analysis suggested the extraction of seven factors. I iteratively removed items that loaded onto their own factor, next attempting a four-factor solution. After dropping another set of items that either created their own factor with one or two items and those that had strong cross-loadings, I was left with 13 items loading on to two factors. An IRT analysis of the remaining items, using a graded-response model and plotting the item-information curves, suggested that 5 items were not providing much additional information and so, for the sake of conciseness were dropped, leading to an eight-item scale, which parallel analysis suggested loaded on to two factors of four items each (TLI = .98, RMSEA = .04). The two factors correlated with each other $r = .41$ and cumulatively explained 61% of the variance of the items. The final items and factor loadings can be seen in Table 4, and the information function curves for the two subscales can be seen in Figure 4. Reliabilities for the scales was acceptable: Effort subscale alpha = .82 [.80, .84], Strategy subscale alpha = .67 [.63, .71].

Table 4
Final Items and Factor Loadings for Final Exploratory Factor Analysis, Study 2

Question	Effort Factor	Strategy Factor
If people just work hard, they can get what they want.	**.88**	-.04
If people work hard enough, they can be whatever they want to be in life.	**.76**	.07
In order to overcome a challenge, all you have to do is try your best.	**.65**	.09
If someone does not achieve their goal, it is because they did not try hard enough.	**.65**	-.08
When someone is stuck on something, it is important to find a new approach.	-.06	**.79**
If someone does well on a challenging task, it must be because they found a way to do it that works well for them.	.11	**.57**
I keep track of my process and, if necessary, I change my techniques or strategies.	.11	**.52**
The best way to complete a challenge must differ from one challenge to another	.06	**.42**

Note: All data come from the exploratory sample of Study 2.


Effort Subscale
Test Information Function
Information
0
2
4
6
8
-4
-2
0
2
4
Ability

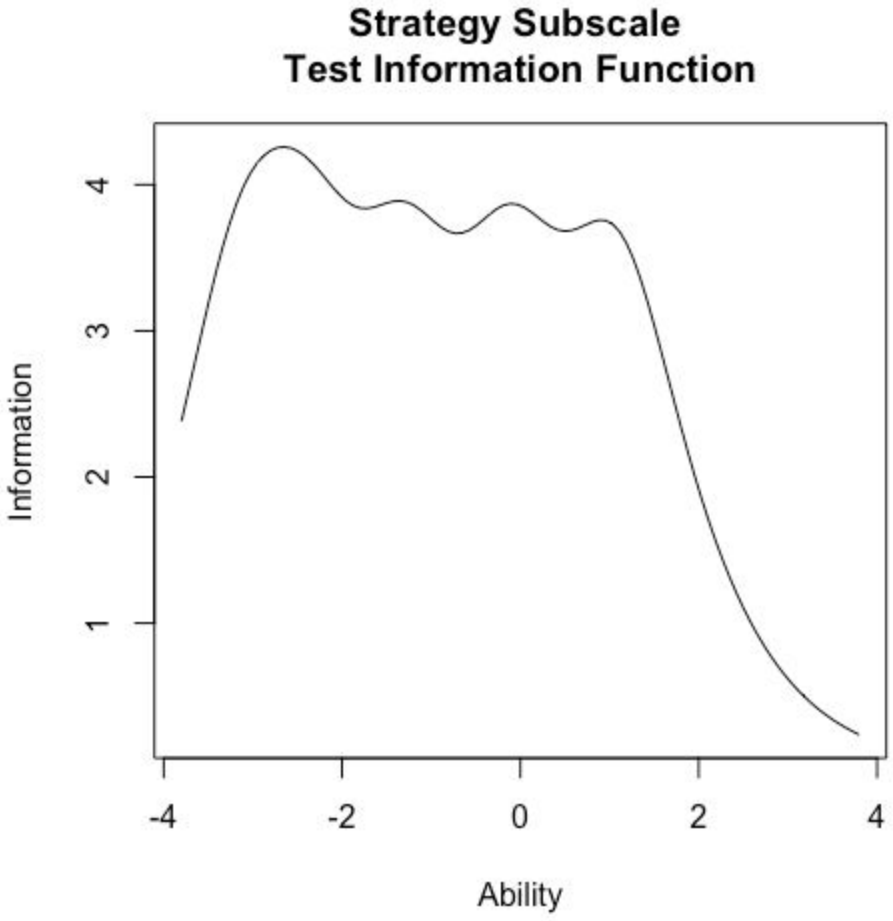


Figure 4. Test Information Curves for the two subscales. Test information curves provide information about where, in the continuum of possible scores, a test is able to discriminate.

Confirmatory Factor Analysis

To validate the structure of the scale, I used three separate samples, the confirmatory sample of the adult collection, the high-school sample, and the college Pretest sample, fitting a two-factor model, allowing the two latent factors to freely covary in each. Fit for this model was good in all samples: nationally-representative adult: CFI = .978, RMSEA = .059 [.043, .076]; high-school: CFI = .965, RMSEA = .058 [.025, .089]; college Pretest: CFI = .953, RMSEA = .047 [.031, .063]. Reliability for the subscales was also relatively adequate in each sample: adult:

Effort subscale alpha = .86 [.84, .87], Strategy subscale alpha = .74 [.71, .77]; high-school sample: Effort subscale alpha = .76 [.72, .81], Strategy subscale alpha = .73 [.68, .78]; college Pretest sample: Effort subscale alpha = .77 [.74, .79], Strategy subscale alpha = .49 [.44, .54].

See Figure 5 for the path diagrams.

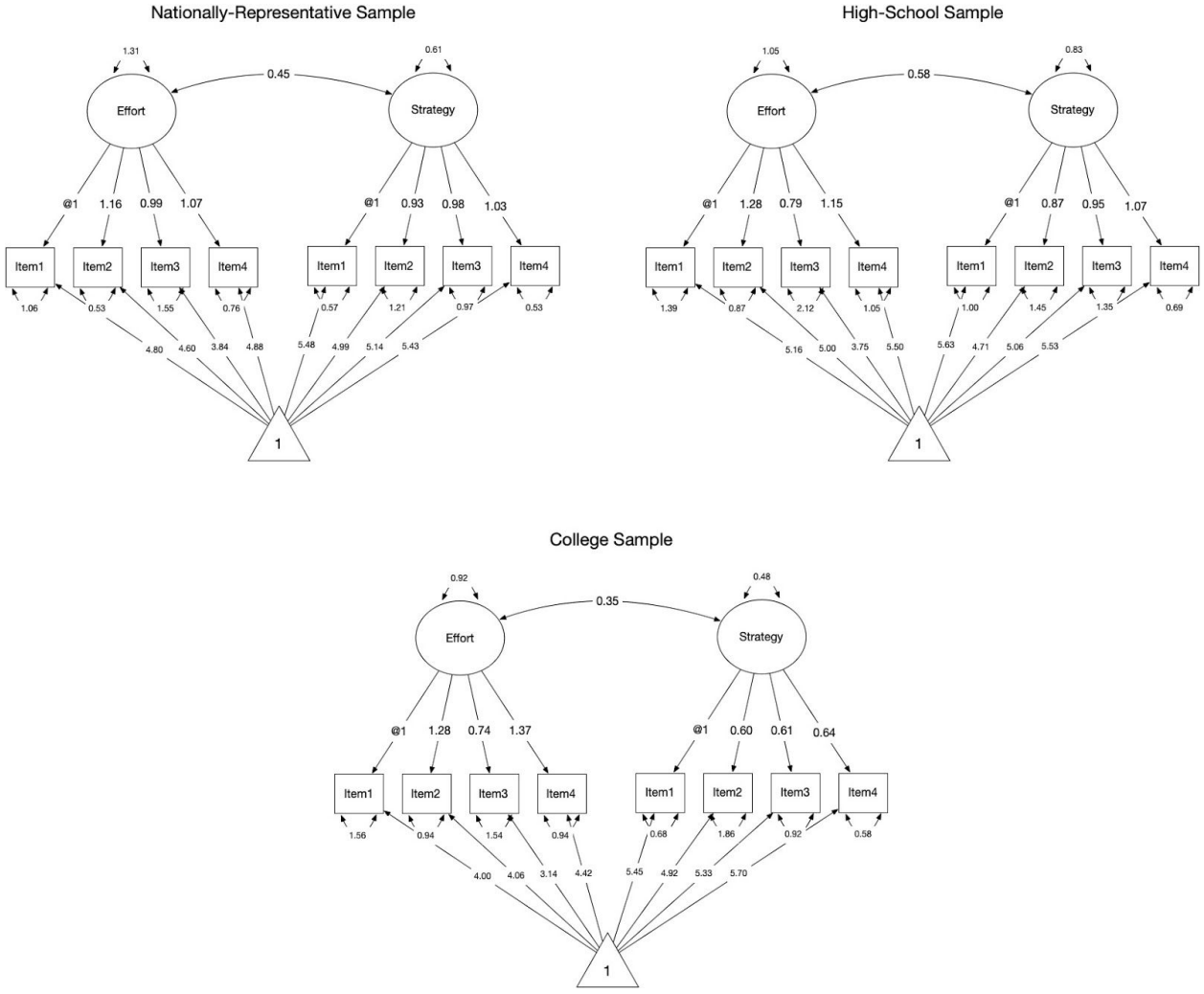


Figure 5. Path diagrams for confirmatory factor analyses. Item E1 = "In order to overcome a challenge, all you have to do is try your best;" Item E2 = "If people just work hard, they can get what they want;" Item E3 = "If someone does not achieve their goal, it is because they did not try hard enough;" Item E4 = "If people work hard enough, they can be whatever they want to be in life;" Item S1 = "If someone does well on a challenging task, it must be because they found a way to do it that works well for them;" Item S2 = "The best way to complete a challenge must

differ from one challenge to another;" Item S3 = "I keep track of my process and, if necessary, I change my techniques or strategies;" Item S4 = "When someone is stuck on something, it is important to find a new approach"

Measurement Invariance

Gender. I found evidence for configural invariance, CFI = .980, RMSEA = .040 [.027, .053], ΔCFI = -.0018; evidence for metric invariance, CFI = .975, RMSEA = .040 [.029, .052], ΔCFI = .0042; marginal evidence for scalar invariance, CFI = .954, RMSEA = .051 [.041, .061], ΔCFI = .021; and evidence for full invariance, CFI = .948, RMSEA = .049 [.040, .059], ΔCFI = .0058.

In analyzing the misfit in scalar invariance, I found that model-estimated mean levels of answers to the prompt: "If someone does not achieve their goal, it is because they did not try hard enough" (Item 3 in the loading on the Effort subscale) had the largest difference across gender. Freeing the mean of that item to differ across men and women led to acceptable fit for partial scalar invariance models, CFI = .963, RMSEA = .046 [.036, .057], ΔCFI = .013; and leaving that one parameter unconstrained additionally led to acceptable fit for partial full-invariance models, CFI = .958, RMSEA = .045 [.035, .055], ΔCFI = .0053; variant parameter: Male = 4.15, *se* = 0.083; Female = 3.68, *se* = 0.066.

Age. I found evidence for configural invariance, CFI = .979, RMSEA = .040 [.027, .053], ΔCFI = -.0014; marginal evidence for metric invariance, CFI = .968, RMSEA = .046 [.034, .057], ΔCFI = .011; marginal evidence for scalar invariance, CFI = .946, RMSEA = .055 [.045, .065] ΔCFI = .023; and evidence for full invariance, CFI = .937, RMSEA = .055 [.045, .064], ΔCFI = .0087.

In analyzing the invariance misfit, I found that relaxing the equality constraint for the loading of the item "If someone does not achieve their goal, it is because they did not try hard enough" (Item 3 in the loading on the Effort subscale) led to acceptable partial metric invariance, CFI = .974, RMSEA = .042 [.030, .054], Δ CFI = .0055. To get to acceptable partial scalar invariance, I also had to relax the constraint on equality between the means of that item, which lead to acceptable fit, CFI = .962, RMSEA = .047 [.036, .057], ΔCFI = .011. No further constraint-releases were needed to get to partial full invariance, CFI = .953, RMSEA = .048 [.038, .058], ΔCFI = .0096; variant parameters: loadings, Young = 1.09, *se* = 0.065; Old = 0.80, *se* = .066; means, Young = 4.05, *se* = 0.73, Old = 3.55, *se* = 0.071.

Life Stage. I found evidence for configural invariance, CFI = .970, RMSEA = .031 [.025, .037], Δ CFI = -.0013; marginal evidence for metric invariance, CFI = .958, RMSEA = .033 [.028, .039], ΔCFI = .013; no evidence for scalar invariance, CFI = .889, RMSEA = .049 [.044, .054] ΔCFI = .068; and no evidence for full invariance, CFI = .839, RMSEA = .053 [.049, .057], ΔCFI = .050.

In analyzing the misfit for the metric invariance, I relaxed the equality of loadings for two items in the college-sample, freeing the loadings for the items "If someone does not achieve their goal, it is because they did not try hard enough" and "If someone does well on a challenging task, it must be because they worked really hard on it" to differ from the loadings for the high-school and adult samples. With these constraints relaxed, I found evidence for partial metric invariance, CFI = .962, RMSEA = .032 [.027, .038], ΔCFI = .0087. To get to acceptable levels of partial scalar invariance, I had to relax another set of constraints, freeing up the means of every effort-scale indicator in the college sample to differ from the other two, and additionally freeing

the mean of the item “If people just work hard, they can get what they want” in the high-school sample to differ from the other two. With those relaxations, I find acceptable evidence for partial scalar invariance, CFI = .942, RMSEA = .037 [.032, .042], ΔCFI = .0197. Finally, to get to partial full invariance, I had to relax the equality constraint on the residual variance of the items “The best way to complete a challenging task is by trying it over and over,” “People's outcomes are determined by their own actions,” and “If someone does not achieve their goal, it is because they did not try hard enough” for the high-school students, and “In order to overcome a challenge, all you have to do is try your best,” “If people just work hard, they can get what they want,” and “If someone does well on a challenging task, it must be because they found a way to do it that works well for them” for the adults. With these constraints relaxed, I found acceptable evidence for partial full invariance: CFI = .934, RMSEA = .036 [.031, .041], ΔCFI = .0078. See Figure 6 for the final path diagram.

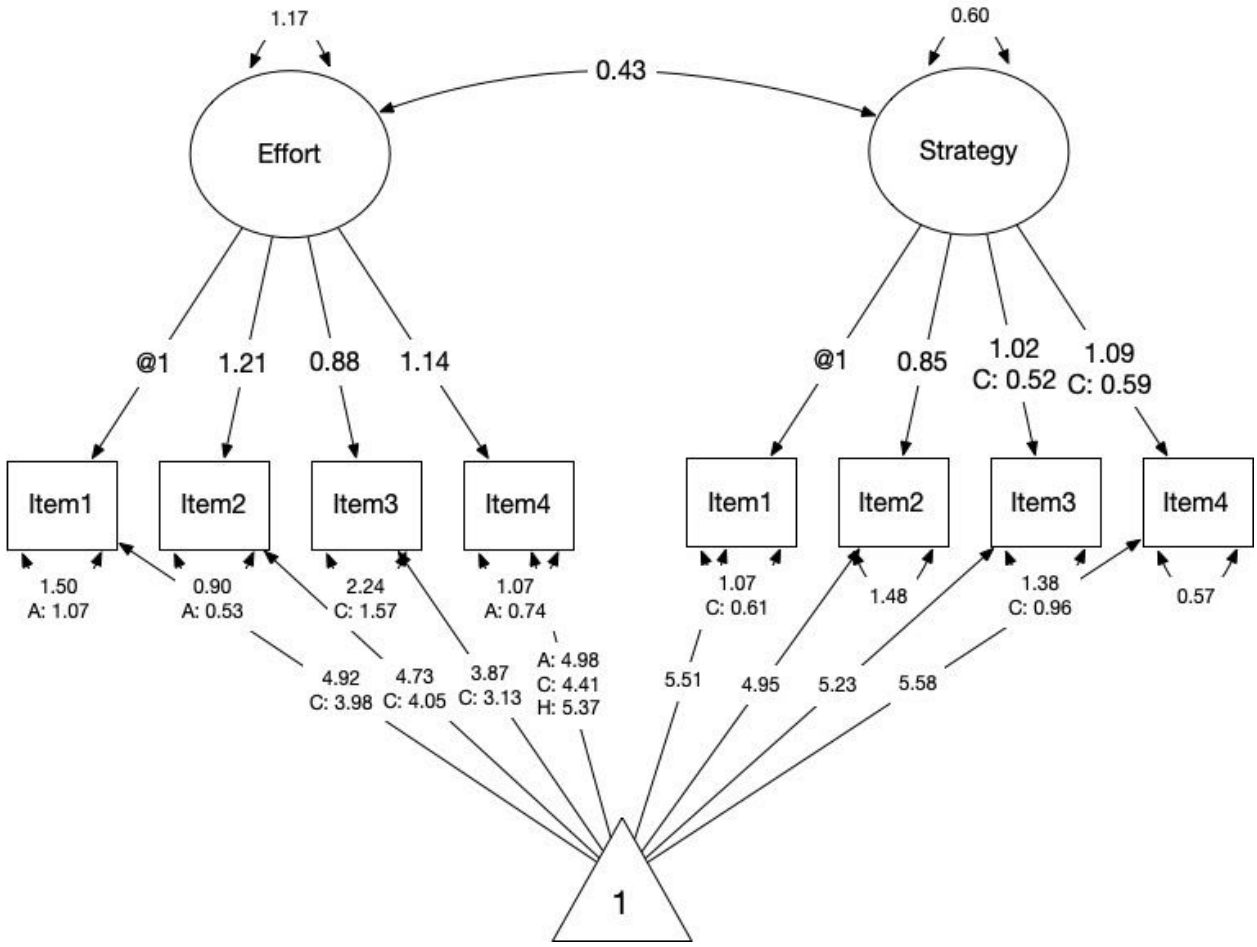


Figure 6. Path diagram for partial measurement invariance across life stage. Values with labels indicate differences in that group across models. A = adult sample; C = college sample; HS = high-school sample. Item E1 = "In order to overcome a challenge, all you have to do is try your best;" Item E2 = "If people just work hard, they can get what they want;" Item E3 = "If someone does not achieve their goal, it is because they did not try hard enough;" Item E4 = "If people work hard enough, they can be whatever they want to be in life;" Item S1 = "If someone does well on a challenging task, it must be because they found a way to do it that works well for them;" Item S2 = "The best way to complete a challenge must differ from one challenge to another;" Item S3 = "I keep track of my process and, if necessary, I change my techniques or strategies;" Item S4 = "When someone is stuck on something, it is important to find a new approach"

Test-Retest Reliability

Scale scores correlated $r(120) = .57$ [.44, .68], $p < .001$ across the two timepoints, ICC = .71 [.63, .77], indicating moderate-to-good reliability over two months. Analyzing the two

subscales separately, it appears that the Effort subscale, $r(120) = .70$ [.59, .78], $p < .001$, ICC = .80 [.74, .84], time 1 alpha = .77 [.74, .79], time 2 alpha = .83 [.79, .86] had better reliability than the Strategy subscale, $r(120) = .42$ [.26, .55], $p < .001$, ICC = .59 [.47, .67], time 1 alpha = .49 [.44, .54], time 2 alpha = .53 [.43, .63].

Discussion

Using three separate samples (total $n = 4{,}106$), including one large nationally representative sample, one sample of high-schoolers, and a college convenience sample, I developed a two-factor, eight-item scale that has relatively-good psychometric properties across a wide variety of potential participants. The two-subscale factor structure fits well in all three samples and the subscales are generally reliable in each. The designed scale has good coverage of the full latent sample-space, allowing for discrimination between respondents at a wide range of possible levels (being able to tell someone who is very high on a dimension from just merely high, and able to tell someone who is very low on a dimension from just merely low). The scale seems to index a relatively-stable trait, as test-retest reliability is surprisingly-good across a two-month timescale, far longer than the usual two-week period.

Perhaps the most impressive psychometric properties of the scale, however, is its remarkable invariance across demographic and life-stage differences. I found evidence for metric invariance between men and women, marginal evidence for invariance between the old and the young in my adult sample, and evidence for metric invariance between my adult and high-school samples (with the college sample, in this case, providing the source of noninvariance). This metric invariance is important - it allows us to reasonably compare mean levels of the construct

of interest (i.e. false growth mindsets) across groups. Simulation studies demonstrate that models with metric invariance but without scalar invariance tend to still perform well in analyses that relate the construct to other phenomena: using a metrically invariant but not a scalar-invariant measurement model as a predictor in a regression or multiple regression, as a mediator, or as a moderator models does not lead to meaningful differences in outcomes as compared to using a fully-invariant model (Guenole & Brown, 2014). There may still, however, be some error if the goal is to identify mean-level differences across groups. Even in that case, however, the error that arises from just one scalar-invariant item, as is the case in the age and in gender models, and in the comparison between high-schoolers and adults, is still relatively minor (Steinmetz, 2013, see also Chen, 2008). It therefore seems reasonable to conclude that the scale can be used in adult populations and with high-schoolers, without having to worry that the values from the scale are functions of a different interpretational process across these groups.

The one group of participants who seemed to be taking the scale differently were the ones participating as part of a college Departmental Pretest. While they clustered items in broadly-similar ways as the adults and high-schoolers (i.e. configural invariance), the relationships between those clustered items differed, especially in the Strategy subscale (i.e. metric non-invariance), and they had a notably different use of the scale points in the Effort subscale (i.e. scalar non-invariance). They were also the group for which subscale reliability was the lowest; the only group with a subscale alpha lower than .70.

There are at least two possibilities as to why this may have been the case. On the one hand, college students, especially those early into their first semester of college (as is the majority of this sample) may be interpreting the items differently than high-schoolers or adults.

There may be something unique about this critical period in my participants' life that makes them especially sensitive to questions of achievement, or especially worried about their ability to make it in this largely novel environment, as researchers have suggested that this period is especially challenging in one's developmental trajectory (e.g. Arnett, 2000; Vaidya et al., 2002).

On the other hand, the differences between the college sample and the adults and high-schoolers may be a function of the method of administration. While all participants took the scale online, those in the college sample participated as part of a Departmental Pretest. In their experimental session, they were asked to complete 31 different psychological scales with nearly 200 separate items combined. This may have been an overwhelming situation for some, which may have led to extra noise in the data, potentially obscuring the relationship between items. The markedly-low reliabilities for the subscales in this sample, especially compared to the reliabilities in the other two samples, may be a function of this scale-fatigue, as may be the 40% of participants in the sample who simply did not complete the scale.

Given the nosiness of the college Pretest sample, it is still somewhat surprising how well the scale held up in a test-retest paradigm, especially given the unusually-long lag between the two timepoints. While two weeks is commonly used as the time interval in reliability testing (see e.g. Deyo, Diehr, & Patrick, 1991), due to sampling limitations I ended up with a much longer two-month interval between test and re-test - a timescale right at the outer bound recommended for retest analyses (Cattell, Eber, & Tatsuoka, 1970). This longer interval, covering as it does important events in students' academic lives, such as their first college mid-term exams and much of the adaptation that comes with their first semester in college, may have led them to reconsider certain assumptions about how best to achieve success or what success even entails

(the so-called "response shift," e.g. Rapkin & Schwartz, 2004; Sprangers & Schwartz, 1999; see also Chmielewski & Watson, 2009 for a discussion of the transient error that comes with measurement over time), which may have added even more noise into the measurement of reliability. Therefore, we can reasonably think about the current value, ICC = .71, as almost a lower-bound to the reliability of the scale, and future measurements in a less chaotic environment (i.e. minimal testing load at the beginning, and a shorter interval between measurements, devoid of major life events), looking at both the reliability of the scale within session (test *dependability*) as well as reliability of the scale over time (test *stability)* may together provide a clearer picture of the stability of the construct over time (see Revelle & Condon, 2019, Watson, 2004; and see Chmielewski & Watson, 2019 for an example on how to estimate test stability by using test dependability to correct for transient error).

Of course, just because a scale has good psychometric properties doesn't mean it's a good scale - in addition to measuring well, it also needs to be measuring *something.* A psychometrically-valid form that's devoid of content is of no meaningful use (see e.g. Maul, 2017), and so, in the next study, I go about situating the construct measured by the false growth mindset scale amongst a set of other constructs, mapping out the psychological space that the scale inhabits.

Study 3: Scale Content

Once a scale has been developed, it should be appropriately situated within a nomological net of related concepts (e.g. Cronbach & Meehl, 1955; Loevinger, 1957) - scores from the scale should relate to scores on conceptually-related scales (convergent validity), it should be shown to

provide new information uncaptured by those related scales (divergent validity) so as to avoid simply restating an earlier concept by a different name (the so-called jangle fallacy; Kelley, 1927), and it should meaningfully predict outcomes in the world (e.g. Campbell & Fiske, 1959).

In order to create a total false growth mindset score, I needed to integrate information from the two subscales. In forming an overall scale value, I chose to combine the Effort and Strategy subscales into a difference score, by simply subtracting average scores on the Strategy subscale from average scores on the Effort subscale, creating a composite that ranges from -6 to + 6. As I conceptualize the false growth mindset as a balance between thinking about the importance of effort and the importance of strategy-use, this difference-score approach allows me to cover the majority of the possible response quadrants, while still being relatively straightforward to calculate and use. In this scoring, low values on the composite indicate an overreliance on the importance of strategy use without effort, while high values indicate an overreliance on effort alone without thinking much about strategies. In choosing to calculate a difference score, however, I give up some understanding of the middle range of the composite, as one could get scores in the region of 0 either from believing highly in both the importance of effort and strategies, or from believing that neither are important. As I anticipate that the latter, more nihilist case is relatively unlikely to occur in my sample, I feel that this is a trade-off worth making. It is, of course, an empirical question as to whether calculating the difference score provides more information than the subscales themselves, and so I test this assumption as well.

In situating the false growth mindset among related constructs, I chose a set of measures that index beliefs about the importance of effort and persistence (the Protestant Work Ethic and grit); about an individual's ability to get things done, under their own control, and their

responsibility for their own individual outcomes (self-efficacy, internal locus of control, and free-will-responsibility); and about their blind optimism towards the future (dispositional optimism). With these constructs, I can tie the false growth mindset to a set of beliefs about the importance of repeated, individual effort, effort that is bound to pay off with no need for outside help, and for the idea that one is solely responsible for one's successes and one's failures - all aspects that orbit the overemphasis on effort, at the expense of strategy-selection, that characterizes the false growth mindset.

To distinguish the false growth mindset scale from other incrementalist beliefs, I also measured beliefs in personal changeability and beliefs in one's ability to grow one's intelligence. Finally, to investigate potential consequences of holding a false growth mindset, I included two measures of societal victim-blaming: just-world beliefs and support for meritocracy, which both inherently assume that those who have bad outcomes in life are themselves responsible for their lower place in society.

Method

Participants

I used the same sampling firm as Study 2 to recruit an additional large non-probability sample of Americans, along the same guidelines as in the exploratory/confirmatory sample of Study 2. In total, 3,118 people participated in the survey, of which 2,323 agreed to have their data used in analysis (M age = 51.51, SD = 11.77; 67.8% female).

Methods

Participants completed the 8-item scale developed in Study 1 (Effort subscale alpha = .86 [.85, .87], Strategy subscale alpha = .80, [.78, .81], correlation between the two subscales $r(2321) = .55$ [.52, .57], $p < .001$) which I formed into a single False Growth Mindset score by simply subtracting the Strategy subscale from the Effort subscale, M = -0.65, SD = 1.11 (possible range from -6 to 6). In addition, participants completed 13 scales to capture a range of related and unrelated constructs: *locus of control* (Sapp & Harrod , 1993), sample item: "My life is determined by my own actions"; *self-efficacy* (Chen, Gully, & Eden, 2001), sample item: "I will be able to achieve most of the goals that I set for myself"; *belief in free will* (Paulhus & Carey, 2011), sample item: "People have complete control over the decisions they make"; the *Protestant Work Ethic* (Katz & Haas, 1988), sample item: "A distaste for hard work usually reflects a weakness of character"; *grit* (Duckworth, Peterson, Matthews, & Kelly, 2007), sample item: "Setbacks don't discourage me"; *belief in a just world* (Reich & Wang, 2015), sample item: "I feel that people who meet with misfortune have brought it on themselves"; *support for meritocracy* (Horberg, Kraus, & Keltner, 2013), sample item: "Society should be structured so that people who are successful, competent or accomplished gain social status and power"; *implicit person theory* (Levy, Stroessner, & Dweck, 1998), sample item: "The kind of person someone is, is something basic about them, and it can't be changed very much"; *growth mindsets about intelligence* (Hong et al., 1999), sample item: "You have a certain amount of intelligence, and you really can't do much to change it"; *dispositional optimism* (Scheier, Carver, & Bridges, 1994), sample item: "In uncertain times, I usually expect the best"; *presence of meaning in life* (Steger, Frazier, Oishi, & Kaler, 2006), sample item: "My life has a clear sense of purpose";

psychological richness (Oishi et al., 2019), sample item: “On my deathbed, I am likely to say ‘I had an interesting life’”; and the *Ten-Item Personality Inventory* (Gosling, Rentfrow, & Swann, 2003).

Scale descriptives for each can be found in Table 5. In addition, participants reported their gender, race, self-assessed place on a socioeconomic status ladder, income, parental education, own education, and political affiliation.

Table 5
Scale Descriptives, Study 3

Scale	**N of Items**	**Alpha**	**Mean (SD)**
Locus of Control	3	.79 [.78, .81]	4.99 (1.13)
Self-Efficacy	8	.95 [.95, .96]	5.27 (1.09)
Free Will	7	.88 [.87, .89]	5.25 (1.07)
Protestant Work Ethic	11	.88 [.88, .89]	4.54 (1.03)
Grit	6	.86 [.85, .87]	5.36 (1.02)
Just World Beliefs	7	.92 [.92, .93]	4.10 (1.31)
Support for Meritocracy	4	.76 [.78, .79]	4.55 (1.17)
Implicit Person Theory	8	.81 [.79, .82]	4.46 (0.98)
Growth Mindset	3	.94 [.93, .94]	4.18 (1.64)
Optimism	6	.81 [.79, .82]	4.44 (1.15)
Meaning in Life	5	.86 [.85, .87]	4.87 (1.32)
Psychological Richness	12	.94 [.94, .95]	4.79 (1.24)
TIPI - Agreeableness	2	.17 [.10, .23]	5.18 (1.16)
TIPI - Conscientiousness	2	.33 [.28, .38]	5.52 (1.21)
TIPI - Extraversion	2	.42 [.37, .47]	3.61 (1.50)
TIPI - Neuroticism	2	.45 [.41, .49]	3.31 (1.34)
TIPI - Openness	2	.04 [-.04, .11]	4.59 (1.16)

Note: All scales scored on a 1 = strongly disagree to 7 = strongly agree scale.

Analytic Strategy

I take a three-pronged, exploratory, approach to understanding the contents of the psychological construct indexed by the false growth mindset scale. In track 1, I investigate whether the scale relates to other theoretically-similar constructs, while making sure that it does not relate to every possible construct. For these analyses, I use simple correlations (Holm-corrected for multiple tests) to assess the degree of relatedness between the various scales.

In track 2, I go one step further, looking to see if the false growth mindset scale can predict patterns of responding across the various scales. Here, I selected only those scales that I expected to be related to the construct, dropping those that I included in the first track as foils to make sure that the scale did not correlate with every possible construct (i.e. psychological richness, personality, basic demographics). I then conducted a latent profile analysis on these scales, creating models that fit a number of possible classes, from models with two classes all the way up to models with ten classes. I selected the best-fitting solution by inspecting both the adjusted Bayesian Information Criterion (aBIC) and the proportion of the sample in the smallest class, using 10% as my cutoff (using the same criteria as in Study 2, above). After creating the classes, I then tested whether one's scores on the false growth mindset scale predicted one's membership in a latent class, i.e. whether the false growth mindset scale predicted one's pattern of responding to the other scale items. To do so, I created a set of logistic regressions, creating a set of dummy variables for class-membership, and then predicting the dummy from one's false growth mindset score.

In track 3, I tested whether the false growth mindset scale could predict outcomes above and beyond other related scales; whether knowing a person's false growth mindset score provides any additional information beyond other related constructs. I chose just world beliefs and support for meritocracy as outcomes of interest as these scales ask about how respondents think about success and failure in the broadest sense - how to judge those who have achieved more or less in their lifetimes, and whether lifetime failure is something that can be seen to be a function of something inherent to a person or whether the structures of society have something to contribute to the selection of who 'wins' or 'loses.' Demonstrating that the false growth mindset scale can predict these outcomes above and beyond other constructs tapping into views about the processes governing the outcomes of others such as the Protestant Work Ethic, which measures how people attribute success and failure to hard work and laziness; free-will beliefs, which measure how much one believes that a person is in control of their own outcomes; or implicit person theory and growth mindsets about intelligence, which measure beliefs about the fixity of personal essences, would be a strong test of the theoretical placement of my construct.

In traditional analyses of incremental predictive validity, one would try to predict an outcome in a multiple-regression framework, with the various competing scales each entered as predictors. If the scale of interest still significantly predicts the outcome of interest when competing scales are included in the model, then one would conclude that the scale has incremental predictive validity for that outcome. However, recent work has demonstrated that using multiple regression to test incremental predictive validity is prone to an inflated false-positive rate, especially when sample sizes are large and the scales in question are of moderate reliability, due to the problem of residual confounding (see e.g. Buttrick, Axt,

Ebersole, & Huband, 2020; Wang & Eastwick, 2020; Westfall & Yarkoni, 2016). Multiple regression assumes that each predictor is perfectly measured with no error, and that therefore in this approach, any correlation between the measurement error in the predictor and in the outcome is incorrectly assumed to be evidence of a relationship between the predictor itself and the outcome itself.

To get around this issue, I use structural equation modelling, an approach which can explicitly model measurement error and therefore take it into account when assessing the relationship between variables. To test for incremental predictive validity, I set up a set of nested models. In the base model, I create measurement models for each scale and outcome, with latent variables identified by their scale items. Each latent predictor is allowed to covary with every other predictor, and I model regression pathways from each predictor to the outcome of interest. To test whether the false growth mindset scale has unique predictive validity, I then remove the regression path from the latent false growth mindset variable to the outcome, and test whether removing this path leads to a significant drop in model fit. If it does, then I can conclude that the regression path is providing new information - that the false growth mindset scale has incremental predictive validity for that outcome; while if the restricted model does not decrease in fit then I can conclude that there is no new information being provided, and therefore that there is no incremental predictive validity for that outcome. See Figure 7 for a schematic path diagram for this type of analysis.

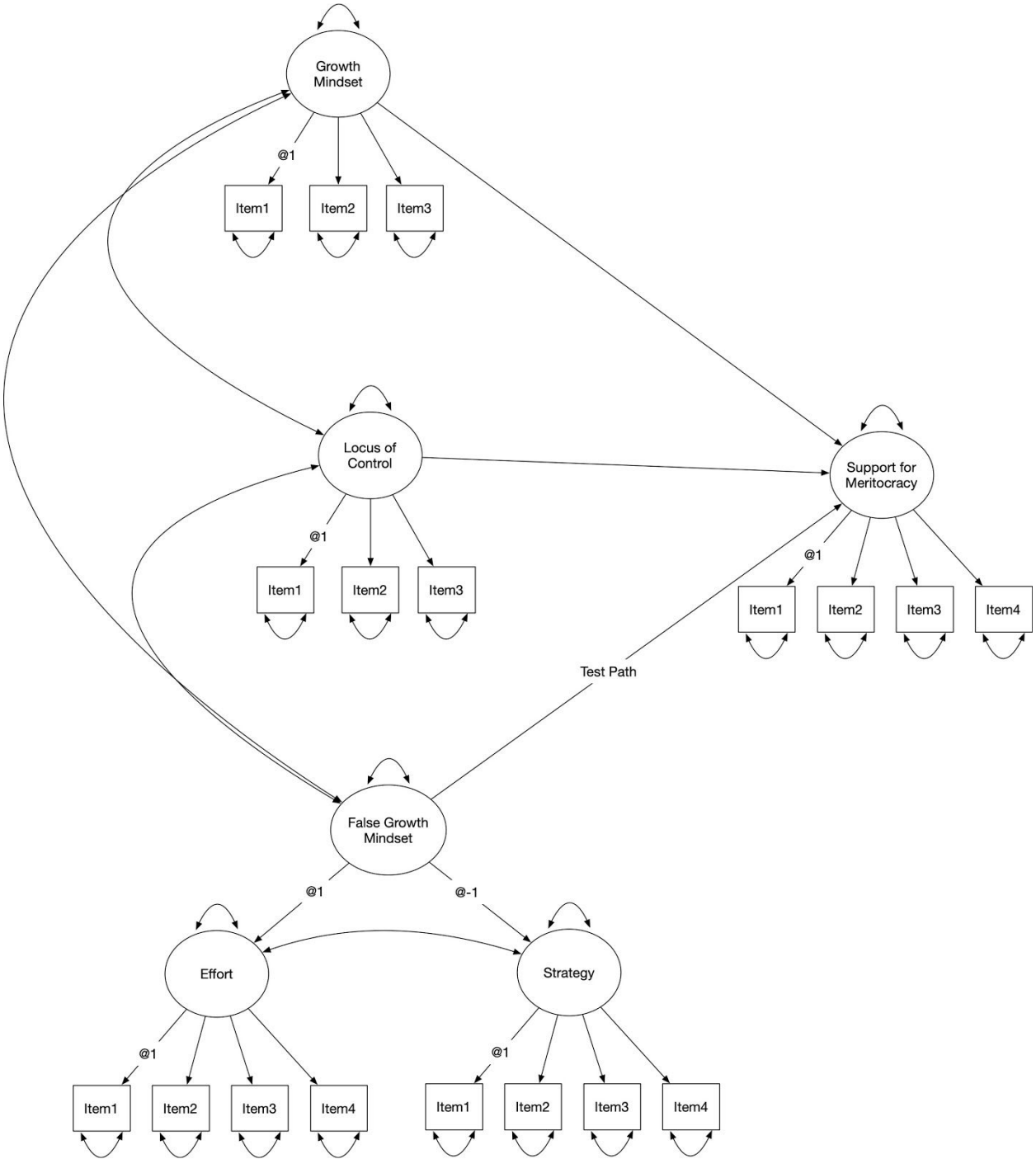


Figure 7. Schematic path diagram for incremental predictive validity analyses. The figure shows just a subset of the predictors for ease of interpretation, and additionally does not show the paths estimating means for every manifest variable. The path removed to test incremental validity claims is marked as the 'Test Path.'

Lastly, to address whether the scale should be treated as a difference score of the two subscales, or whether the two subscales should remain independent predictors of various outcomes, I fit an additional set of SEM models, predicting just world beliefs and support for meritocracy from the effort subscale latent variable, from the strategy subscale latent variable, and from the combined false growth mindset latent variable. If the combined latent variable provides any information above and beyond the subscales themselves, then removing that pathway, in a set of nested models, should lead to a decrease in model fit. See Figure 8 for a schematic path diagram

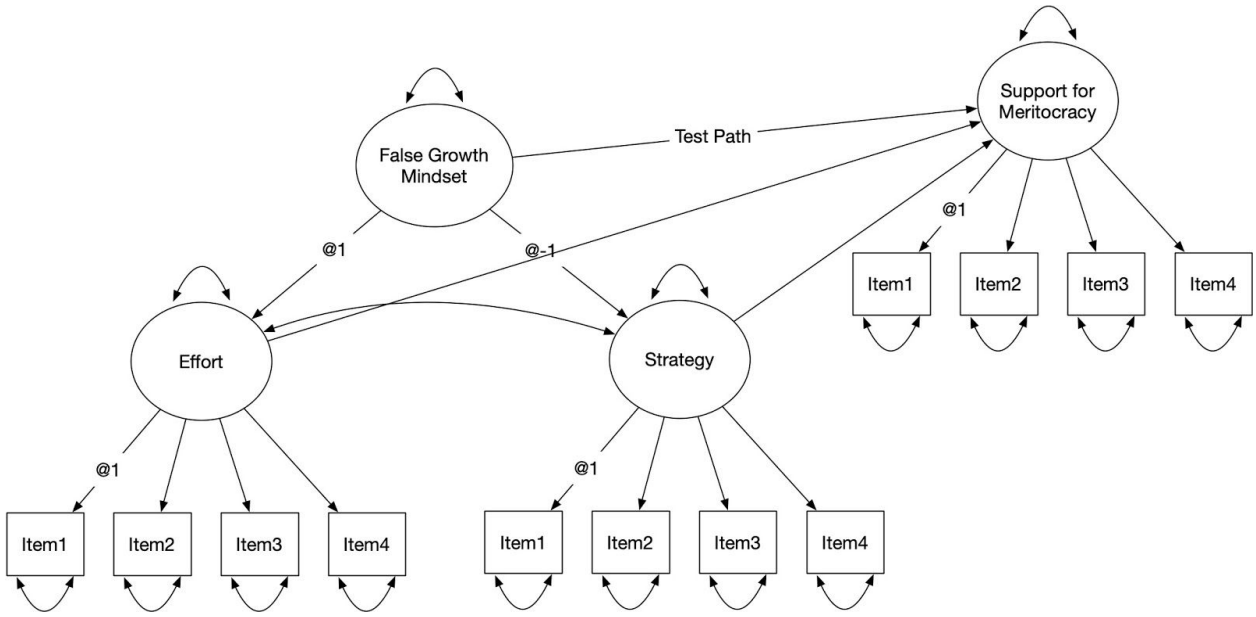


Figure 8. Schematic path diagram for subscale scoring analyses. The figure shows just a subset of the construct predictors for ease of interpretation, and additionally does not show the paths estimating means for every manifest variable. The path removed to test incremental validity claims is marked as the 'Test Path.'

Results

Distribution of Subscale Values

I looked at the distributional properties of the two subscales, dividing responses into four quadrants using the subscale midpoints: high Effort, high Strategy (at or above the midpoint on both scales); low Effort, low Strategy (below the midpoint on both scales), high Effort, low Strategy; and low Effort, high Strategy. As anticipated, relatively few participants were categorized as low Effort, low Strategy, n = 68 (2.9%). The majority of participants were categorized as high Effort, high Strategy, n = 1,694 (72.9%); followed by low Effort, high Strategy, n = 530 (22.8%); and high Effort, low Strategy, n = 31 (1.3%). See Figure 9.

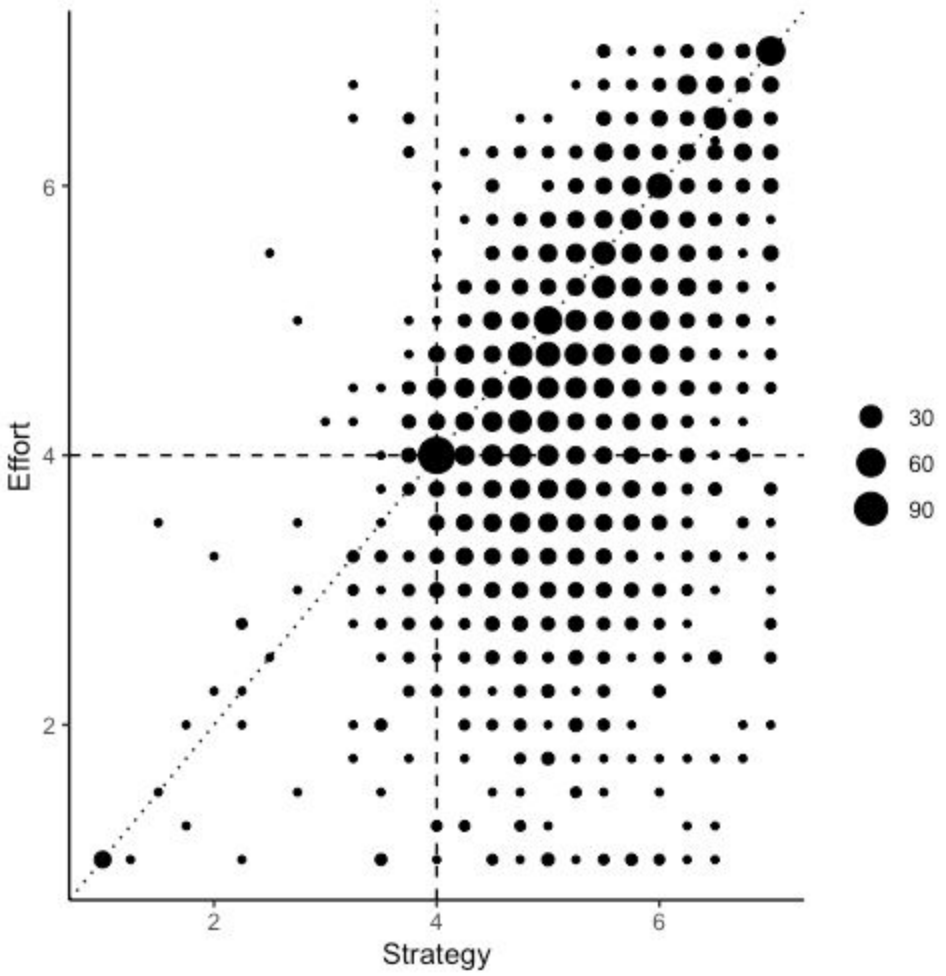


Figure 9. Plot of responses to the two subscales. Circle size represents the number of people at each point. The dashed lines indicate the midpoints of the two subscales, and the dotted line indicates the midpoint of the difference-score scale. Circles above the dotted line represent positive false growth mindset scores (i.e. Effort greater than Strategy), and circles below the dotted line indicate negative false growth mindset scores (i.e. Strategy greater than Effort).

Convergent Validity

I found that the False Growth Mindset scale correlated meaningfully with Just World beliefs, $r(2,321) = .47$ [.44, .50], $p < .001$; with the Protestant Work Ethic, $r(2,320) = .35$ [.32, .39], $p < .001$; with support for meritocracy, $r(2,320) = .27$ [.24, .31], $p < .001$; with Free Will beliefs, $r(2,321) = .27$ [.23, .31], $p < .001$; and with one's internal locus of control, $r(2,321) = .26$

[.22, .29], $p < .001$. The scale correlated more weakly with one's self-reported place on the SES ladder, $r(2,321) = .19$ [.15, .23], $p < .001$; with Self-Efficacy beliefs, $r(2,321) = .11$, $p < .001$; and with dispositional optimism, $r(2,321) = .083$ [.042, .12], $p < .001$.

The scale correlated negatively with growth mindset beliefs $r(2,319) = -.13$ [-.17, -.093]; and did not correlate with implicit person theory, $r(2,320) = -.036$ [-.076, .005], $p = .084$. Surprisingly, the scale also did not correlate with political orientation, $r(2,312) = .04$ [-.001, .08], $p = .055$, or with grit, $r(2,323) = 0.03$ [-.01, .07], $p = .12$. For all correlations with the False Growth Mindset Measure, see Table 6, and for the full set of correlations between measures, see Figure 10. See Supplement for the correlations between each of the False Growth Mindset Scale subscales with the various other measures.

Table 6
Correlations with False Growth Mindset and Measures, Study 3

Measure	**n**	***r* [95% CI]**	***p*-value**
Locus of Control	2323	0.26 [0.22, 0.29]	< .001
Self-Efficacy	2323	0.11 [0.07, 0.15]	< .001
Free Will	2323	0.27 [0.23, 0.31]	< .001
Protestant Work Ethic	2322	0.35 [0.32, 0.39]	< .001
Grit	2323	0.03 [-0.01, 0.07]	.12
Just-World Beliefs	2323	0.47 [0.44, 0.5]	< .001
Meritocracy Support	2322	0.27 [0.24, 0.31]	< .001
Implicit Person Theory	2322	-0.04 [-0.08, 0.00]	.08
Growth Mindset	2321	-0.13 [-0.17, -0.09]	< .001
Optimism	2323	0.08 [0.04, 0.12]	< .001
Meaning in Life	2323	0.09 [0.05, 0.13]	< .001
Richness	2322	0.04 [0.00, 0.08]	.04
Extraversion	2321	0.08 [0.04, 0.12]	< .001
Agreeableness	2322	-0.15 [-0.19, -0.11]	< .001
Conscientiousness	2322	-0.16 [-0.20, -0.12]	< .001
Neuroticism	2322	-0.01 [-0.05, 0.03]	.55
Openness	2322	-0.15 [-0.19, -0.11]	< .001
Age	2322	-0.15 [-0.19, -0.11]	< .001
Gender	2321	-0.03 [-0.07, 0.01]	.21
SES Ladder	2323	0.19 [0.15, 0.23]	< .001
Income	2322	0.08 [0.04, 0.12]	< .001
Education (Self)	2321	-0.05 [-0.09, 0.00]	.03
Education (Parents)	2316	0.02 [-0.02, 0.06]	.32
Politics	2314	0.04 [0.00, 0.08]	.06

Note. Gender is scored 1 = Male, 2 = Female, 3 = Non-Binary. Politics is scored 1 = Very Liberal to 7 = Very Conservative.

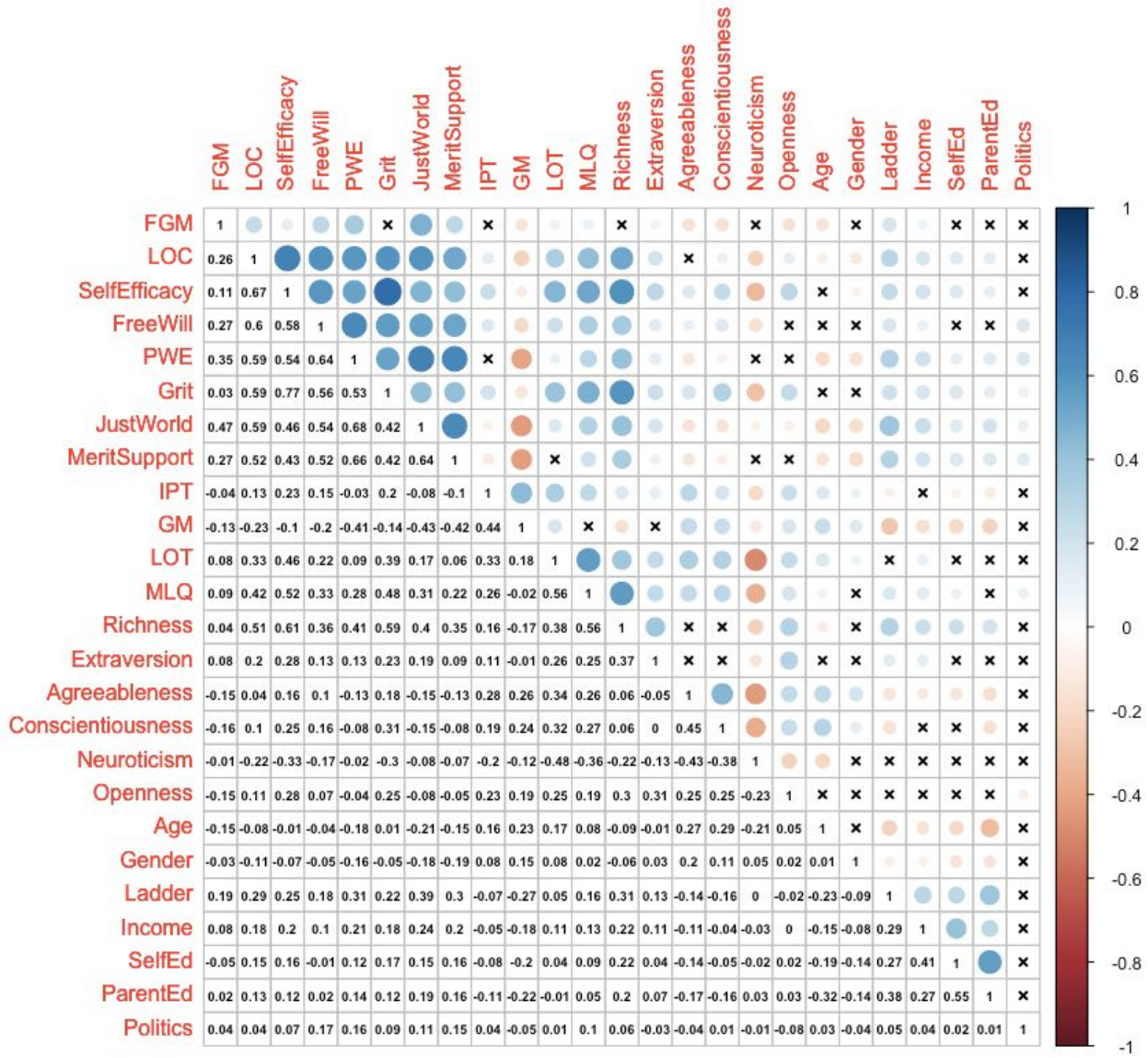


Figure 10. Correlations between all measures in Study 2. Coefficients are printed below the diagonal, and all correlations with *p*-values > .05 (holm-corrected for multiple tests) are indicated with an X above the diagonal. FGM = False Growth Mindset, LOC = Locus of Control, PWE = Protestant Work Ethic, IPT = Implicit Person Theory, GM = Growth Mindset, LOT = Dispositional Optimism, MLQ = Meaning in Life.

Predicting Profile Membership

I constructed profiles from the locus-of-control, self-efficacy, Protestant Work Ethic, support for meritocracy, just-world, grit, implicit person theory, growth mindset, free-will, and dispositional optimism scales, along with political orientation. I fit a set of potential profiles, from ones enforcing two classes, all the way up to ones enforcing ten classes. Based on aBIC and the proportion of people in each profile, I ended up selecting a three-profile solution, with 768 people in Profile 1 (33.9% of respondents), 1,195 people in Profile 2 (51.6% of respondents), and 336 people in Profile 3 (14.5% of respondents). See Table 7 for statistics for each profile, and see Figure 11 for the three-profile solution.

Table 7
Fit Statistics for Latent Profile Analyses, Study 3

# of Profiles	Parameters	LL	AIC	BIC	aBIC	Entropy	Smallest Profile Proportion
2	34	-37928.459	75924.919	76120.352	76012.327	.854	.34
3	46	-36554.722	73201.443	73465.853	73319.702	.859	.15
4	58	-35786.909	71689.818	72023.203	71838.926	.885	.027
5	70	-35371.19	70882.38	71284.742	71062.339	.864	.018
6	82	-34940.386	70044.772	70516.11	70255.58	.838	.019
7	94	-34707.527	69603.054	70143.369	69844.713	.852	.017
8	106	-34460.888	69133.775	69743.066	69406.284	.857	.0099
9	118	-34314.022	68864.044	69542.311	69167.403	.857	.0099
10	130	-34209.754	68679.509	69426.753	69013.718	.835	.0099

Note: LL = Log Likelihood; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; aBIC = Sample-Size-Adjusted Bayesian Information Criterion

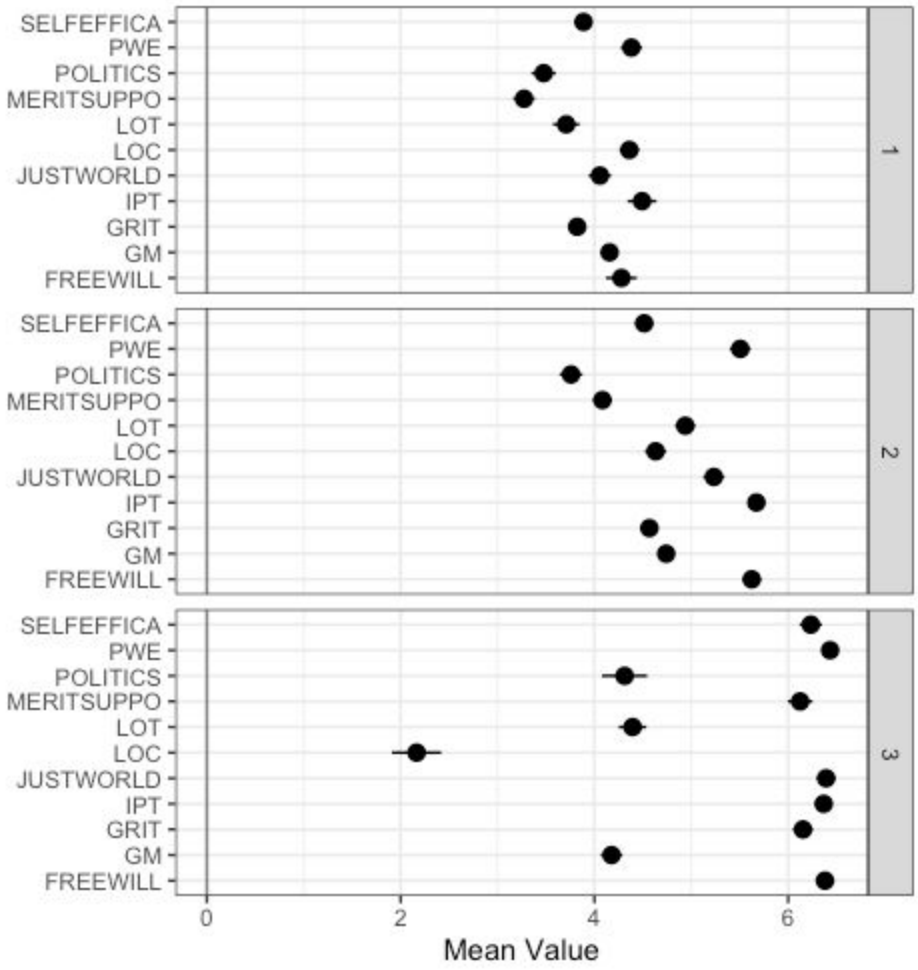


Figure 11. Three-profile latent profile analysis solution. PWE = Protestant Work Ethic; LOT = Dispositional optimism; LOC = Locus of control; IPT = Implicit person theory; GM = Growth mindset. Error bars indicate 95% CIs

In inspecting the profiles, it appears that one of them, Profile 3, best fits the conceptualization of the False Growth Mindset: high levels of self-efficacy, high levels of the Protestant Work Ethic, high support for meritocracy, high levels of just-world beliefs, high levels of grit, high levels of free-will beliefs, high levels of belief in personal change, and moderately

conservative political beliefs. Interestingly, this group also has a relatively high degree of entity theorizing about intelligence.

I then analyzed whether one's false growth mindset score predicted membership in Profile 3. I found that it did (n = 2,317): OR = 1.077 [1.063, 1.090], $X^2(1) = 132.35$, $p < .001$, McFadden's pseudo-R^2 = .074. A first-quartile false growth mindset score predicted a 10.1% chance of belonging to this profile, while a median score predicted a 15.7% chance of belonging to this profile, and a third-quartile score predicted a 19.3% chance of belonging to this profile (33% higher than the base-rate of membership in this profile, which was 14.5% of the total sample).

This relationship with the false growth mindset scale was unique to Profile 3, as false growth mindset negatively predicted membership in Profile 1, (n = 2,317): OR = 0.916 [0.901, 0.932], $X^2(1) = 101.72$, $p < .001$, pseudo-R^2 = .032, first-quartile predicted probability = 39.1%, median = 32.6%, third-quartile = 28.2%, base-rate = 33.9%; and did not predict membership in Profile 2, (n = 2,317): OR = 1.014 [0.995, 1.033], $X^2(1) = 2.16$, p = .14, pseudo-R^2 = .00064, first-quartile predicted probability = 50.8%, median = 51.8%, third-quartile = 52.5%, base-rate = 51.6%. I also find broadly similar results in the next-best fitting set of profiles, the six-factor solution. See the Supplement for details.

Incremental Validity

I used two sets of nested SEM analysis to test the incremental predictive validity of the false growth mindset scale on both just world beliefs and support for meritocracy, above and beyond political affiliation, locus of control, self-efficacy, the Protestant Work Ethic, grit, implicit person theory, growth mindset, dispositional optimism, and free-will-beliefs. In the

model predicting just world beliefs, I found evidence for a significant loss of model fit when removing the regression pathway from the latent false growth mindset variable to the latent just world belief variable, indicating that the false growth mindset scale does have incremental predictive validity for this construct: $X^2(1) = 2{,}663.59$, $p < .001$. I found similar evidence for the incremental predictive validity of the false growth mindset scale for support for meritocracy: $X^2(1) = 26{,}055.25$, $p < .001$. See Table 8 for fit statistics for the two SEM analyses, and see Table 9 for regression parameter estimates for the SEM analyses. For the matching regression parameter estimates from the multiple-regression analyses, see Table S3 in the Supplement.

Table 8
Fit Statistics for Incremental Validity Analyses, Study 3

Model	Parameters	df	-2LL	AIC	CFI	RMSEA	Δ -2LL	*p*-value
Just World Beliefs								
Full	260	157628	477080.6	161824.6	.736	.079 [.078, .080]	-	-
Restricted	259	157628	479744.1	164486.1	.713	.082 [.081, .083]	2663.59	< .001
Meritocracy Support								
Full	251	150671	456501.4	155159.4	.740	.079 [.076, .079]	-	-
Restricted	250	150672	482556.6	181212.6	.499	.11 [.108, .110]	26055.25	< .001

Note: -2LL = -2 Log Likelihood; AIC = Akaike Information Criterion; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Table 9
Structural Regression Parameters for Incremental Validity Analyses, Study 3

	Just-World Beliefs	**Meritocracy Support**
Scale	***b***	***b***
False Growth Mindset	1.38	17.84
Locus of Control	23.97	-0.56
Self-Efficacy	15.07	0.14
Free-Will	-12.18	-0.36
Protestant Work Ethic	5.42	-4.09
Grit	-19.20	5.30
Implicit Person Theory	-13.76	2.50
Growth Mindset of Intelligence	-5.89	0.64
Dispositional Optimism	-4.01	-1.38
Political Orientation	1.06	0.11

Note: Political Orientation is scored 1 = Very Liberal to 7 = Very Conservative. *p*-values for regression parameters are not calculated by default, and the model had trouble calculating standard errors or 95% CIs for all regression parameters.

Subscale Scoring

Using two sets of nested models, I found that treating the false growth mindset scale as a difference score between the effort and strategy subscales provided additional information in predicting just world beliefs above and beyond the two subscales themselves, as removing that pathway led to a significant decrease in model fit, $X^2(1) = 1{,}319.34$, $p < .001$; while treating the scale as a difference score did not provide additional information for predicting support for meritocracy, $X^2(1) = 0.00$, $p = 1.00$. See Table 10 for fit statistics for the two SEM analyses.

Table 10
Fit Statistics for Incremental Subscale Analyses, Study 3

Model	Parameters	df	-2LL	AIC	CFI	RMSEA	Δ -2LL	*p*-value
Just World Beliefs								
Full	50	34777	104764.0	35210.00	.948	.075 [.070, .079]	-	-
Restricted	49	34778	106083.3	36527.34	.887	.11 [.106, .114]	1319.34	< .001
Meritocracy Support								
Full	41	27820	86508.31	30868.31	.931	.087 [.080, .093]	-	-
Restricted	40	27821	86508.31	30868.31	.931	.086 [.080, .092]	0.00	1.00

Note: -2LL = -2 Log Likelihood; AIC = Akaike Information Criterion; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Discussion

Using a large nationally-representative sample, situating the false growth mindset scale among a set of conceptually-related scales, I find that the false growth mindset scale fits neatly into the proposed nomological net. The scale correlates with many of the scales that it has been theorized to relate to, such as the Protestant Work Ethic, locus of control, self-efficacy, and free-will beliefs, while being clearly distinct from prior measures of belief about the changeability of persons (no correlation) and growth mindsets about intelligence (an inverse correlation). In addition to correlating with single items, the false growth mindset scale also predicts the theorized pattern of responding, with those high in a false growth mindset more likely to jointly have high levels of self-efficacy, belief in the Protestant Work Ethic, support for meritocracy, just-world beliefs, grit, the changeability of persons, and free-will beliefs, while being more politically-conservative and with moderate entity-theories about intelligence - almost precisely the profile I anticipated finding. Finally, I find that the false growth mindset scale does

not simply tap into already-existing scales, as it predicts levels of societal victim-blaming, via just-world beliefs and support for meritocracy, above and beyond all competing scales in my dataset.

However, since all my data is self-reported, I do not yet have any evidence that the scale is related to any pragmatically-relevant behaviors. It is entirely possible that the construct measured by the false growth mindset scale simply does not affect behavior in any meaningful way. It may also be the case that getting people to change their false growth mindset beliefs does not, by itself, lead to any downstream consequences of note. I have, in other words, no evidence yet of the causal power of the false growth mindset. Future directions should be aimed at this gap in the current studies, looking to manipulate variables that should change one's beliefs about how go about striving for success and then measuring how that change leads to changes in actual behavior, ensuring that the false growth mindset scale can act as a mediator between a theoretically relevant independent variable and a theoretically relevant dependent variable.

There are at least two promising directions that such research can take: looking at how the false growth mindset affects perceptions of others, and how it affects perceptions of the self. When it comes to perceptions of others, a false growth mindset may have its strongest impact in how one interprets their failures. Given the strong relationships between the false growth mindset scale and measures of societal victim-blaming, such as just world beliefs and support for meritocracy, messages that target false growth mindset beliefs directly, either by increasing them, such as by reinforcing the importance of effort in doing well at things, or by decreasing them, such as by reinforcing the importance of asking for help and trying different sorts of strategies in different sorts of situations, should change the way that people view those who have

failed to achieve success, with manipulations that increase the false growth mindset leading to higher levels of victim-blaming and stigmatization and manipulations that decrease the false growth mindset leading to lower levels of victim-blaming and stigmatization.

When it comes to perceiving the self, the false growth mindset may affect how people view their own abilities after being confronted with repeated failure. For those with a true growth mindset, failure is seen as a call to try harder, but also to change up one's strategies and to seek out help. When one has tried one's hardest, in other words, there's still the possibility to improve by trying 'smarter.' Failure after one has put in the effort is still frustrating - even those with a true growth mindset become anxious when their performance fails to improve after they've put in the effort (Plaks & Stecher, 2007), but this must be vastly more distressing to those with a false growth mindset, who only have one element in their improvement toolkit: trying harder still. If it seems as though trying harder is getting them nowhere, those with a false growth mindset may have no other attributional option than to start blaming themselves, and may come to see their abilities as limited. After all, if they've tried their hardest and still not succeeded, then they may just not have what it takes in that particular domain. In the face of sustained challenge, they may, in other words, become entity theorists.

General Discussion

Across three studies, sampling responses from over 10,000 participants, I find evidence that the false growth mindset, a belief that anyone can succeed as long as they try hard enough, is a real, widespread construct, that it relates to pragmatically-important outcomes, that it can be

reliably measured as a construct, and that the construct is kin, but not identical, to other measures of growth mindsets, predicting societal victim-blaming above and beyond competing measures.

In Study 1, I find, using a nationally-representative sample of 9th graders and their mathematics teachers, that teachers who believe in an extreme growth mindset and who tend to engage in one-size-fits-all authoritarian effort-praise, those, in other words, holding a false growth mindset, make up a substantial proportion of the general teacher population, and that students in their classrooms see them, ironically, as holding fixed mindsets. Their students also tend themselves towards a more fixed view of their own intelligence, and these two beliefs that their students hold, both about their own abilities and about the way they perceive that their teacher thinks about their students, lead to decreased performance in school.

In Studies 2 and 3, I use a number of samples, largely drawn from nationally-representative pools, to design and validate a direct measure of the false growth mindset, looking at how people think about the importance of both effort and strategies for success. I find that the scale is psychometrically-solid, that it performs relatively similarly across a number of different populations, and that it has at least adequate test-retest reliability, suggesting that it indexes a trait-like attitude towards the world. The construct itself is related to a number of similar constructs, such as the Protestant Work Ethic and beliefs about individual responsibility for one's actions, is only loosely related to other individual incrementalist views of human nature, but does predict a pattern of belief within individuals - relative political conservatism and near-ceiling levels of agreement with the Protestant Work Ethic, belief in self-efficacy, support for meritocracy, just-word beliefs, free-will beliefs, grit, and the generalizable changeability of people (though not when it comes to intelligence, where this

group tends towards a more essentialized view) - that matches my understanding of what the construct entails. Finally, I show that false growth mindsets are not just the sum total of these constructs, demonstrating that the scale predicts societal victim-blaming, in the form of just-world beliefs and support for meritocracy, above and beyond all competing measures.

Thanks to my sampling plan, I feel fairly confident that the bulk of this work generalizes to most American contexts. My somewhat limited college sample may restrict the generality of the failure to extend measurement invariance to this group, and to the finding of only moderate test-retest reliability for the scale, but I would predict that the teacher and student sampling of Study 1 and the content analyses of Study 3 should generalize to other high-school settings and to other American adults, respectively.

I am less confident, however, about whether my results will generalize to a non-American context. While growth mindsets themselves appear to be adaptive in cross-cultural settings (see Walton & Wilson, 2018 for a table of successful international implementations of growth mindset interventions), the false growth mindset, grounded as it is in an individualistic understanding of merit, where anyone can succeed as long as they try hard enough, may be a uniquely American worldview. While American psychology may not be unique, it is certainly distinctive, having been shaped by a history of frontiership and a relatively high level of residential mobility (e.g. Kitayama, Conway, Pietromonaco, Park, & Plaut, 2010; Oishi, 2010). Thanks to these historical and socioecological trends, Americans are more likely to be individualistic, more likely to believe that the future will be better than the past, and more likely to believe in their own ability to get things done by themselves and be properly rewarded for those efforts. American belief in meritocracy, in other words, may be unusually high (see, e.g.

Alesina, di Tella, & MacCullough, 2004; Ladd & Bowman, 1988). Given these background worldviews, it may be that the false growth mindset is stronger and more coherent in the United States, and that similar analyses in more egalitarian or collectivist/responsibilist settings may not show the same sorts of prevalence or factor structure for the construct, and therefore a false growth mindset, where it exists, may have different effects on people's behaviors and outcomes.

In addition to the cross-cultural validity of the false growth mindset construct, an additional open question is whether the construct is domain general or more encapsulated: whether a false growth mindset is something that applies to all aspects of one's goal pursuits; or whether it can be active in the perception of some domains, such as education, while not present in others, such as addiction, obesity, or other occasions of weakness of will. The present work demonstrates that beliefs about the changeability of people generally is not identical with beliefs about the changeability of intelligence, as the two constructs only correlate $r = .44$, a finding that echoes other researchers, who demonstrate that incremental theories are often domain-specific (e.g. Dweck, Chiu, & Hong, 1995; Yeager & Dweck, 2012, Ziegler & Stoeger, 2010). However, seeing as the false growth mindset is a relatively abstract concept, it is still an open question as to whether it acts as a general background assumption about how the world works, or whether people can hold beliefs about effort that only apply to education, say, but not to obesity, and therefore, whether it is possible to shift general false growth mindsets and not shift beliefs in more specific domains.

Similarly, it is still an open question whether, at the general level, there is one false growth mindset that applies equally strongly to beliefs about the self and to interpretations of others; or whether people are more willing to apply false growth mindset beliefs to others while

reserving more charitable beliefs for themselves. Psychology is rife with self/other asymmetries, especially when it comes to self-serving attributions, or attributing one's own successes to internal merit and one's own misfires to external causes while failing to extend that same benefit of the doubt to others (e.g. Miller & Ross, 1975; Pronin, Lin, & Ross, 2002). Given this common belief-structure, it may be that people also play by different rules when judging the importance of effort and attributing blame for themselves and for others, focusing on the importance of effort for others, but focusing, say, on the importance of the situation for one's self. At the same time, there is evidence that people lean on their theories of their own malleability when judging the behavior of others, with those with more incremental theories more likely to blame others for their failures (e.g. Ryazanov & Christenfeld, 2018), which may suggest that those who believe that their own efforts are all that are needed for success extend that belief when thinking about the outcomes of others.

A third open question is in addressing the potential positive benefits of holding a false growth mindset. I suspect that there are situations where holding a false growth mindset is actually more beneficial than holding a true growth mindset or a purely entity theory - situations in which the key to success really is trying harder and continuing to persist. As a false growth mindset can be conceptualized, in part, as an extremely high belief in one's self-efficacy and grit, those with a false growth mindset may persist at a task longer than those with other mindsets, which, in situations where pure persistence is rewarded, should increase success. Similarly, those with a false growth mindset may also be more likely to take on initial challenges, heedless of the potential difficulty, as they may believe that they can do anything as long as they work hard at it. Of course, I hypothesize that these beliefs become maladaptive if the task is, in fact, too hard, or

if it does not give way to hard work alone, and in these situations a person with a false growth mindset may find themselves wasting their time in perseveration at a fruitless task, or may come to decide that, in fact, they do not have what it takes in that domain, and may find their belief in their own efficacy shattered into a more negative entity theorizing.

A final important future direction would be to better understand how a person comes to the worldview underlying the false growth mindset. I have suggested two possible routes: one route may come from over-emphasis of effort in growth-mindset training materials; and another route may comes from a complex of background beliefs about effort and meritocracy (and some surely comes from a combination of both, whereby those who already hold the underlying worldview read the intervention materials in a more motivated way, picking up on the elements of the growth mindset that agree with their pre-existing sense of the world). There may be other pathways as well. Previous work has looked at the effects of outcome praise in the development of an entity theory of intelligence (e.g. Mueller & Dweck, 1988), and it may be that similar processes can build a false growth mindset - if one constantly praises a person's effort, they may come believe that effort is, in fact, the only thing that matters, inadvertently sensitizing them into a false-growth-mindset way of approaching tasks.

In conclusion, I provide evidence that there is something called a false growth mindset that differs from the currently-understood concepts of incremental lay theories or growth mindsets about intelligence. This difference comes with consequences, and those that hold a false growth mindset have different ideas about success and failure than those with a true growth mindset, and are viewed differently by outside observers. This difference may help to explain part of the mystery of the heterogeneity of mindset effects. By lumping those with true growth

mindset along with those with false growth mindsets, researchers may be inadvertently combining two different belief structures, which may then lead to erratic and noisy point estimates, both for assessing the outcomes of growth mindset interventions (e.g. Sisk et al., 2018), where those that take away a false growth mindset may have different learning outcomes than those with a true growth mindset; and for assessing the relationship between mindset beliefs and psychological outcomes such as challenge-seeking and performance after failure feedback (e.g. Burgoyne et al., 2020), where those that have a false growth mindset may have a different motivational orientation than those with a true growth mindset. By considering the false growth mindset as a novel kind, we may be able to get a better picture of the effects of the true growth mindset, and how to best ensure that people are able to reap the benefits of the mindset, while minimizing the potential downsides.

References

Alesina, A., Di Tella, R., & MacCulloch, R. (2004). Inequality and happiness: are Europeans and Americans different? *Journal of Public Economics, 88*(9-10), 2009-2042.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*(4), 556-574

Amemiya, J., & Wang, M. T. (2018). Why effort praise can backfire in adolescence. *Child Development Perspectives, 12*(3), 199-203.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist, 55*(5), 469–480.

Bergman, L. R. & Trost, K. (2006). The person-oriented versus the variable-oriented approach: are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Quarterly, 52*, 601–632.

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.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis, 39*(1), 146-170.

Browman, A. S., Miele, D. B., O'Dwyer, L. M., & May, S. C. (under review). *Teachers' growth mindsets and the differential treatment of high- and low-ability students.*

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, N.Y.: Guilford Press.

Burnette, J. (in prep). *A call for improving implementation fidelity: The case of growth mindset interventions*. North Carolina State University, Raleigh, N.C.

Burnette, J. L., Forsyth, R. B., Desmarais, S. L., & Hoyt, C. L. (2019). Mindsets of addiction: Implications for treatment intentions. *Journal of Social and Clinical Psychology, 38*(5), 367-394.

Burnette, J. L., Hoyt, C. L., Dweck, C. S., & Auster‐Gussman, L. (2017). Weight beliefs and messages: Mindsets predict body‐shame and anti‐fat attitudes via attributions. *Journal of Applied Social Psychology, 47*(11), 616-624.

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139*(3), 655-701

Burgoyne, A. P., Hambrick, D. Z., & Macnamara, B. N. (2020). How firm are the foundations of mind-set theory? The claims appear stronger than the evidence. *Psychological Science, 31*(3), 258-267.

Butler, R. (1994). Teacher communications and student interpretations: Effects of teacher responses to failing students on attributional inferences in two age groups. *British Journal of Educational Psychology, 64*(2), 277-294

Buttrick, N., Axt, J., Ebersole, C. R., & Huband, J. (2020). Re-assessing the incremental predictive validity of Implicit Association Tests. *Journal of Experimental Social Psychology, 88*, 103941.

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

Campbell, D. T., Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances, 5*(2), eaau4734.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: Institute for Personality and Ability Testing.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*(1), 62-83.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*(5), 1005–1018.

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Chiu, C. Dweck, C. S., Tong, J. Y., & Fu, J. H. (1997). Implicit theories and conceptions of morality. *Journal of Personality and Social Psychology, 73*(5), 923–40

Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319.

Chmielewski, M., & Watson, D. (2009). What is being assessed and why it matters: The impact of transient error on trait research. *Journal of Personality and Social Psychology, 97*(1), 186-202.

Cozzarelli, C., Wilkinson, A. V., & Tagler, M. J. (2001). Attitudes toward the poor and attributions for poverty. *Journal of Social Issues, 57*, 207–227

Crandall, C. S. (1994). Prejudice against fat people: ideology and self-interest. *Journal of Personality and Social Psychology, 66*(5), 882–894.

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

Deyo, R. A., Diehr, P., & Patrick, D. L. (1991). Reproducibility and responsiveness of health status measures: Statistics and strategies for evaluation. *Controlled Clinical Trials, 12*(4 suppl), 142S–158S.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087-1101.

Dweck, C. S. (1999). *Self-Theories: Their role in motivation, personality and development.* Philadelphia: Taylor and Francis/Psychology Press.

Dweck, C. S. (2006). *Mindset.* New York: Random House

Dweck, C. S. (2008). Can personality be changed? The role of beliefs in personality and change. *Current Directions in Psychological Science, 17*, 391-394.

Dweck, C. S. (2015). Carol Dweck revisits the growth mindset. *Education Week, 35*(5), 20-24

Dweck, C. S. (2016, January 11). Recognizing and overcoming false growth mindset. *Edutopia*. Retrieved from https://www.edutopia.org/blog/recognizing-overcoming-false-growth-mindset-carol-dweck

Dweck, C. S. (2016, January 13). What having a growth mindset actually means. *Harvard Business Review*. Retrieved from: https://hbr.org/2016/01/what-having-a-growth-mindset-actually-means

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry, 6*(4), 267-285.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273.

Dweck, C. S., & Molden, D. C. (2008). Self-Theories: The Construction of Free Will. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are We Free?: Psychology and Free Will* (pp. 44–64). Oxford: Oxford University Press

Ferguson, R. F., Phillips, S. F., Rowley, J. F., & Friedlander, J. W. (2015). *The influence of teaching beyond standardized test scores: Engagement, mindsets, and agency*. Retrieved from The Achievement Gap Initiative at Harvard University website: http://www. agi. harvard. edu/publications.php.

Finch, W. H. & French, B. F. (2014) Multilevel Latent Class Analysis: Parametric and nonparametric models. *The Journal of Experimental Education, 82*(3), 307-333.

Fuchs, L. S., Fuchs, D., & Phillips, N. (1994). The relation between teachers' beliefs about the importance of good student work habits, teacher planning, and student achievement. *The Elementary School Journal, 94*(3), 331-345

Furnham, A. (1984). The Protestant work ethic: A review of the psychological literature. *European Journal of Social Psychology, 14*(1), 87-104

Gaspard, H., Wille, E., Wormington, S. V., & Hulleman, C. S. (2019). How are upper secondary school students' expectancy-value profiles associated with achievement and university STEM major? A cross-domain comparison. *Contemporary Educational Psychology, 58*, 149-162.

Georgiou, S. N., Christou, C., Stavrinides, P., & Panaoura, G. (2002). Teacher attributions of student failure and teacher behavior toward the failing student. *Psychology in the Schools, 39*(5), 583–595

Gervey, B. M., Chiu, C., Hong, Y., & Dweck, C. S. (1999). Differential use of person information in decisions about guilt versus innocence: The role of implicit theories. *Personality and Social Psychology Bulletin, 25*(1), 17–27

Gopalan, M., & Tipton, E. (2018, November 3). *Is the National Study of Learning Mindsets Nationally-Representative?* https://doi.org/10.31234/osf.io/dvmr7

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality, 37*, 504-528.

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology, 5*, 980.

Haimovitz, K., & Dweck, C. S. (2016). Parents' views of failure predict children's fixed and growth intelligence mind-sets. *Psychological Science, 27*(6), 859-869

Haimovitz, K., & Dweck, C. S. (2017). The origins of children's growth and fixed mindsets: New research and a new proposal. *Child Development, 88*(6), 1849-1859.

Hafer, C. L., & Bègue, L. (2005). Experimental research on just-world theory: Problems, developments, and future challenges. *Psychological Bulletin, 131*(1), 128–167

Hafer, C. L., & Olson, J. M. (1989). Beliefs in a just world and reactions to personal deprivation. *Journal of Personality, 57*(4), 799–823

Halberstadt, A. G., Dunsmore, J. C., Bryant, A., Parker, A. E., Beale, K. S., Thompson, J. A., & Thompson, J. A. (2013). Development and validation of the Parents' Beliefs About Children's Emotions Questionnaire. *Psychological Assessment, 25*(4), 1195–1210

Heider, F. (1958). *The psychology of interpersonal relations*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Publishers.

Henry, K. L. & Muthén, B. (2010). Multilevel Latent Class Analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling, 17*(2), 193-215.

Hong, Y. Y., Chiu, C. Y., Dweck, C. S., Lin, D. M. S., & Wan, W. (1999). Implicit theories, attributions, and coping: a meaning system approach. *Journal of Personality and Social Psychology, 77*(3), 588-599.

Horberg, E. J., Kraus, M. W., & Keltner, D. (2013). Pride displays communicate self-interest and support for meritocracy. *Journal of Personality and Social Psychology, 105*(1), 24-37.

Hoyt, C. L., Burnette, J. L., Auster-Gussman, L., Blodorn, A., & Major, B. (2017). The obesity stigma asymmetry model: The indirect and divergent effects of blame and changeability beliefs on antifat prejudice. *Stigma and Health, 2*(1), 53–65.

Hoyt, C. L., Burnette, J. L., Thomas, F. N., & Orvidas, K. (2019). Public health messages and weight-related beliefs: Implications for well-being and stigma. *Frontiers in Psychology, 10*, 2806

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

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88-110

Hussey, I. & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*(2), 131-155

Kammrath, L. K., & Peetz, J. (2012). You promised you'd change: How incremental and entity theorists react to a romantic partner's promised change attempts. *Journal of Experimental Social Psychology, 48*(2), 570-574.

Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology, 55*(6), 893-905.

Kelley, T. L. (1927). *Interpretation of Educational Measurements.* Yonkers-on-Hudson, N.Y.: World Book Company

Kitayama, S., Conway, L. G. III, Pietromonaco, P. R., Park, H., & Plaut, V. C. (2010). Ethos of independence across regions in the United States: The production-adoption model of cultural change. *American Psychologist, 65*(6), 559–574

Koo, T. K. & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163.

Kuznetsova A., Brockhoff P. B., & Christensen R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.

Ladd, E. C., & Bowman, K. H. (1998). *Attitudes toward economic inequality.* Washington, D.C.: AEI Press.

Ledgerwood, A., Mandisodza, A. N., Jost, J. T., & Pohl, M. J. (2011). Working for the system: Motivated defense of meritocratic beliefs. *Social Cognition, 29*(3), 322–340

Lerner M. J. (1980). *The belief in a just world – A fundamental delusion*. Boston, MA: Springer

Leroy, N., Bressoux, P., Sarrazin, P., & Trouilloud, D. (2007). Impact of teachers' implicit theories and perceived pressures on the establishment of an autonomy supportive climate. *European Journal of Psychology of Education, 22*(4), 529-545.

Levy, S. R., & Dweck, C. S. (1999). The impact of children's static versus dynamic conceptions of people on stereotype formation. *Child Development, 70*, 1163–80

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*(6), 1421-1436.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, N.Y.: Guilford Press

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(3), 635–694

Madon, S., Willard, J., Guyll, M., & Scherr, K. C. (2011). Self‐fulfilling prophecies: Mechanisms, power, and links to social problems. *Social and Personality Psychology Compass, 5*(8), 578-590

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*(2), 147-186

Maul, A. (2017). Rethinking traditional methods of survey validation. *Measurement: Interdisciplinary Research and Perspectives, 15*(2), 51-69.

Maxwell, S. E. & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model-Comparison Perspective*. New York, NY: Psychology Press.

McCoy, S. K., & Major, B. (2007). Priming meritocracy and the psychological justification of inequality. *Journal of Experimental Social Psychology, 43*(3), 341-351

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. J*ournal of Applied Psychology, 93*(3), 568-592.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.

Miller, D. T. & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*(2), 213-225.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance.* New York, N.Y.: Routledge.

Monroe, A. E., Dillon, K. D., & Malle, B. F. (2014). Bringing free will down to Earth: People's psychological concept of free will and its role in moral judgment. *Consciousness and Cognition, 27*, 100-108

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75*(1), 33–52.

Muenks, K., Canning, E. A.,Green, D. J., Zirkel, S., Garcia, J. A., & Murphy, M. C. (invited revision). *Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset predict in-class psychological experiences*.

Murrah, W. M., Kosovich, J., & Hulleman, C. (2017). A framework for incorporating intervention fidelity in educational evaluation studies. In G. Roberts, S. Vaughn, S. N. Beretvas, & V. Wong (Eds.), *Treatment fidelity in studies of educational intervention* (pp. 39–60). New York: Routledge.

Murphy, M. C., & Dweck, C. S. (2010). A culture of genius: How an organization's lay theory shapes people's cognition, affect, and behavior. *Personality and Social Psychology Bulletin, 36*(3), 283-296

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*, 33-84.

Oishi, S. (2010). The psychology of residential mobility: Implications for the self, social relationships, and well-being. *Perspectives on Psychological Science, 5*(1), 5-21.

Oishi, S., Choi, H., Buttrick, N., Heintzelman, S. J., Kushlev, K., Westgate, E. C., ... & Ng, B. W. (2019). The psychologically rich life questionnaire. *Journal of Research in Personality, 81*, 257-270.

Park, D., Gunderson, E. A., Tsukayama, E., Levine, S. C., & Beilock, S. L. (2016). Young children's motivational frameworks and math achievement: Relation to teacher-reported instructional practices, but not teacher theory of intelligence. *Journal of Educational Psychology, 108*(3), 300-313.

Paulhus, D. L., & Carey, J. M. (2011). The FAD–Plus: Measuring lay beliefs regarding free will and related constructs. *Journal of Personality Assessment, 93*(1), 96-104.

Pizarro, D. A., & Tannenbaum, D. (2012). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91-108). Washington, DC: American Psychological Association.

Plaks, J. E., Grant, H., & Dweck, C. S. (2005). Violations of implicit theories and the sense of prediction and control: Implications for motivated person perception. *Journal of Personality and Social Psychology, 88*(2), 245–262

Plaks, J. E., & Stecher, K. (2007). Unexpected improvement, decline, and stasis: A prediction confidence perspective on achievement success and failure. *Journal of Personality and Social Psychology, 93*(4), 667–684

Polit, D. F. (2014). Getting serious about test–retest reliability: a critique of retest research and some recommendations. *Quality of Life Research, 23*(6), 1713-1720.

Portney, L. G. & Watkins, M. P. (2015). *Foundations of clinical research: applications to practice*. Philadelphia, P.A.: F. A. Davis.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling, 18*, 161- 182.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209-233.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 28*(3), 369-381.

Putnick, D. L. & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90.

Rapkin, B. D., & Schwartz, C. E. (2004). Towards a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health and Quality of Life Outcomes,* 2(1), 14.

Rattan, A., Savani, K., Komarraju, M., Morrison, M. M., Boggs, C., & Ambady, N. (2018). Meta-lay theories of scientific potential drive underrepresented students' sense of belonging to science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology, 115*(1), 54-75

Reich, D. A., & Arkin, R. M. (2006). Self-doubt, attributions, and the perceived implicit theories of others. *Self and Identity, 5(*2), 89-109

Reich, B., & Wang, X. (2015). And justice for all: Revisiting the global belief in a just world scale. *Personality and Individual Differences, 78*, 68-76.

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment, 31*(12), 1395-1411.

Robins, R. W., & Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity, 1*(4), 313-336

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart and Winston

Rutkowski, L. & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*(1), 31-57.

Ryazanov, A. A., & Christenfeld, N. J. S. (2018). Incremental mindsets and the reduced forgiveness of chronic failures. *Journal of Experimental Social Psychology, 76*, 33–41.

Sapp, S. G., & Harrod, W. J. (1993). Reliability and validity of a brief version of Levenson's locus of control scale. *Psychological Reports, 72*(2), 539-550.

Savani, K., Stephens, N. M., & Markus, H. R. (2011). The unanticipated interpersonal and societal consequences of choice: Victim blaming and reduced support for the public good. *Psychological Science, 22*(6), 795–802

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. Journal of Personality and Social Psychology, 67(6), 1063-1078.

Schneider, W. & Pressley, M. (1989). *Memory development between 2 and 20*. New York: Springer-Verlag

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

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*, 549-571

Smith, A. (2020). *When "good" beliefs go bad: The upsides and downsides of believing that emotions are controllable.* Unpublished Master's Thesis, University of Toronto.

Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality-of-life research: A theoretical model. *Social Science and Medicine, 48*(11), 1507–1515.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The Meaning in Life Questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*, 80-93.

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9*(1), 1-12

Stride C.B., Gardner S., Catley. N. & Thomas, F. (2015). *Mplus code for mediation, moderation, and moderated mediation models*. Retrieved from. http://www.offbeat.group.shef.ac.uk/FIO/mplusmedmod.htm

Sun, K. L. (2019). The role of mathematics teaching in fostering student growth mindset. *Journal for Research in Mathematics Education, 49*(3), 330-355.

Tay, L. & Jebb, A. T. (2018). Establishing construct continua in construct validation: the process of continuum specification. *Advances in Methods and Practices in Psychological Science, 1*(3), 375-388.

Thompson, A. G. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics, 15*(2), 105-127.

Tingley, D. Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5), 1-38.

Tipton, E., Yeager, D. S., Iachan, R., & Schneider, B. (2019). Designing probability samples to study treatment effect heterogeneity. In P. Lavrakas, M. Traugott, C. Kennedy, A. Holbrook, E. de Leeuw, & B. West (Eds.) *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*, (pp. 435-456), Hoboken, NJ: John Wiley & Sons.

Tullett, A. M. & Plaks, J. E. (2016). Testing the link between empathy and lay theories of happiness. *Personality and Social Psychology Bulletin, 42*(11), 1505–1521

Vaidya, J. G., Gray, E. K., Haig, J., & Watson, D. (2002). On the temporal stability of personality: Evidence for differential stability and the role of life experiences. *Journal of Personality and Social Psychology, 83*(6),1469–1484.

van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal, 47*(2), 497-527.

Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review, 125*(5), 617–655

Walton, G. M. & Yeager, D. S. (in press). Seed and soil: Psychological affordances in contexts to explain where wise interventions succeed or fail. *Current Directions in Psychological Science*

Wang, Y. A., & Eastwick, P. W. (2020). Solutions to the problems of incremental validity testing in relationship science. Personal Relationships, 27(1), 156-175.

Watson, D. (2004). Stability versus change, dependability versus error: Issues in the assessment of personality over time. *Journal of Research in Personality, 38*(4), 319-350.

Weiner, B. (1994). Integrating social and personal theories of achievement striving. *Review of Educational Research, 64*(4), 557–573

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum. R. M. (1972). Perceiving the causes of success and failure. In E. E. Jones, D. E. Kanouse, H. K. Kelly, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: perceiving the causes of behavior* (pp. 95-120). Momstown, NJ: General Learning Press

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PloS one, 11*(3). E0152719.

Wormington, S. (under review). I*s effort enough or do I need strategies, too? Assessing the fidelity and efficacy of a growth mindset intervention in developmental mathematics.* University of Virginia

Wormington, S. V., Corpus, J. H., & Anderson, K. G. (2012). A person-centered investigation of academic motivation and its correlates in high school. *Learning and Individual Differences, 22*(4), 429-438.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist, 47*(4), 302-314.

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Paunesku, D. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature, 573*(7774), 364-369.

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... & Trott, J. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*(3), 374-391.

Yeager, D. S., Trzesniewski, K. H., & Dweck, C. S. (2013). An implicit theories of personality intervention reduces adolescent aggression in response to victimization and exclusion. *Child Development, 84*(3), 970-988.

Yeager, D. S., Trzesniewski, K. H., Tirri, K., Nokelainen, P., & Dweck, C. S. (2011). Adolescents' implicit theories predict desire for vengeance after peer conflicts: Correlational and experimental evidence. *Developmental Psychology, 47*(4), 1090-1107.

Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., ... & Gomez, E. M. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences, 113*(24), 3341-3348

Ziegler, A., & Stoeger, H. (2010). A learning oriented subjective action space as an indicator of giftedness. *Psychological Science Quarterly, 50*(2), 222–236.