Title
Perceived Implicit Theories of Intelligence and Academic Help-Seeking

Author
Mehta, Smriti

Publication Date
2023

Perceived Implicit Theories of Intelligence and Academic Help-Seeking

By

Smriti Mehta

A dissertation submitted in partial satisfaction of the

requirement for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Silvia A. Bunge, Chair
Professor Dacher J. Keltner
Professor Leif D. Nelson
Professor Barry J. Schwartz
Professor Mark R. Wilson

Summer 2023

Abstract

Perceived Implicit Theories of Intelligence and Academic Help-Seeking

by

Smriti Mehta

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Silvia A. Bunge, Chair

Implicit theories of intelligence refer to individuals' beliefs about the fundamental nature of intelligence. On one end of this spectrum are entity theorists, who believe intelligence is a static entity that cannot be altered. On the other end are incremental theorists, who believe intelligence can be modified through action. An incremental view of intelligence has been linked to adaptive motivational patterns and better academic outcomes. It stands to reason, however, that adaptive motivational patterns will be beneficial only in educational contexts that are supportive and foster growth. Consequently, research assessing educators' theories about intelligence and how they relate to student outcomes has emerged in recent years.

The research presented in this dissertation suggests that students' perceptions of their instructors' beliefs about intelligence, and behaviors that relay those perceptions, can be reliably measured and are related to a constellation of student cognitions in the classroom—this includes students' belief in their intellectual potential, attitude toward academic help-seeking, and sense of efficacy about achieving academically. Help-seeking behavior, however, was unrelated to self-reported help-seeking behavior across three studies. Results consistently showed that when students perceived their instructors to hold a malleable view of intelligence, they perceived the learning environment as less competitive. This finding is potentially relevant for educational contexts plagued by higher levels of competition. Evidence also suggested that these perceptions are related to higher course engagement, lower concerns about negative evaluation, and lower negative feelings like belonging uncertainty and impostor feelings.

I begin the dissertation by exploring the lay of the theoretical land and motivating the research topic (Chapter 1). Next, in Chapter 2, I provide preliminary evidence that suggests that students' perceptions of their instructors' implicit theories correlate with their attitude toward academic help-seeking and several other sociopsychological outcomes. In Chapter 3, I present evidence that suggests that these theories can be measured as a unidimensional construct in a principled way by describing the development and validation of an instrument designed to measure Perceived implicit Theories of Intelligence (P-TOI). Lastly, in Chapter 4, I report the results of an observational study that links P-TOI with some of the hypothesized variables related to students' psychological experience in rigorous STEM (Science, Technology, Engineering, and Mathematics) courses. I end with a discussion of the theoretical implications and limitations of this work and speculate briefly on future directions.

I dedicate this dissertation to my mother, who was neither allowed to remain in a classroom nor be in front of one as she was meant to be.

Amma, nothing I will ever do shall make me as proud as I am to be your daughter.

Acknowledgments

To my entire department, I am deeply indebted, for the formative experiences of the last four years would not have been possible without their acceptance and support. To be part of a community with such a rich intellectual tradition has been the greatest honor of my life.

Within the department, Dr. Frank Sulloway has played an integral role in my evolution as a scholar and has made my time in bear territory infinitely more enjoyable. The steadfast dedication and creative flair with which he approaches his work is something I can only aspire to. Outside Berkeley, I am immensely grateful to Dr. Daniël Lakens, a kindred spirit whose kindness, friendship, and informal mentorship have been invaluable to me. His unrelenting drive to improve scientific and psychological practice is a constant source of inspiration.

I owe much to them both, including the completion of this dissertation, which would be inconceivable without their selfless support.

To my dissertation committee, I extend the most profound gratitude.

First and foremost, to my advisor, Dr. Silvia Bunge, who has been generous in innumerable ways, and without whom this dissertation would not be possible. She is a guiding model of a rigorous researcher, an intuitive mentor, and a socially responsible scientist.

To Dr. Dacher Keltner, whose appreciation for the wonders of human nature is contagious. His advice to students stands out, in my mind, as the best of its kind: “Know your methods, go to a mystery, and do your homework.”

To Dr. Leif Nelson, who has nudged me to be adequately skeptical, but no less optimistic, about our discipline. I owe him for welcoming me into an academic community I deeply cherish, for the respect and encouragement I have seldom deserved (but always appreciated), and most importantly, for being an intellectual sparring partner bar none.

To Dr. Barry Schwartz, whose infinite wit, wisdom, and knack for expressive clarity are a joy to behold. His vast knowledge and astute understanding of human behavior make every conversation as edifying as it is enjoyable. It is an honor to have him as an official friend and unofficial rabbi.

And finally, to my secondary advisor, Dr. Mark Wilson, who provided me with a much-needed intellectual home and much-appreciated feedback by hand. I am grateful he respected me enough to tell me when I was wrong and graciously adopted me into a community where I could gain the tools to be less so.

To know and learn from each of you has been a tremendous privilege.

Table of Contents

List of Tables

List of Figures

List of Appendices

CHAPTER I

PERCIEVED THEORIES OF INTELLIGENCE AND ACADEMIC HELP-SEEKING

Within psychological literature, the study of motivation is unparalleled in terms of the depth of its history and the richness of its scope. *All* behavior is motivated, one way or another, and human motivation has been studied under various labels—will, volition, instincts, wants, needs, and drives—across the entire history of the discipline by those who have shaped that history (Bandura, 1986; Freud, 1915; James, 1981; Lewin, 1938; Tolman, 1922). Motivation research initially focused on its biological, mechanistic aspects such as motor movement, drive, and energy (Atkinson, 1964; Weiner, 1972, 1980; 1990). This genesis is unsurprising given the etymological roots of 'motivation' in the Latin term *motivus*, meaning 'to move.' For decades, this research domain remained abstract and dominated by behaviorist stimulus-response theories (Atkinson, 1964; Hull, 1943), and not until the 1980s and 90s did achievement motivation, which addresses complex behavior and necessitates a cognitive component, become a mainstay in motivation research.

During the cognitive revolution of the 1950s, 60s, and 70s, the focus in motivation research shifted from behavior to cognition, following the trends of the time (Gardner, 1987; G. A. Miller, 2003). Prior to this shift, the study of motivation had been conflated with learning—specifically, learning that could be measured objectively—since gauging motivation was (and often still is) accomplished through the assessment of learning (Atkinson, 1964; Weiner, 1990). Motivation researchers were especially dissatisfied with the simplistic stimulus-response models that could not adequately explain cognitive influences on behavior (Dember, 1974). Following the transition from behaviorism to cognitive psychology, the study of motivation acquired a *social-cognitive* lens, which emphasized the role of personal agency within sociocultural influences (Bandura, 1977, 1989, 2001; Dweck, 1986; Schunk & DiBenedetto, 2020; Weiner, 1986).

As a result of this paradigm shift, theories about internal factors that affect academic performance—e.g., goal orientation, approach vs. avoidance orientation, causal attribution of academic outcomes—proliferated (Weiner, 1990). It was acknowledged that students' construal and interpretation of academic situations influence their achievement-related behaviors (Dweck & Leggett, 1988). Inspired by leading theories like learned helplessness (Seligman, 1968, 1972) and achievement goal orientation (Dweck, 1986), researchers began to evaluate how different attributions and cognitions lead to different motivational patterns, labeling adaptive motivational patterns as 'mastery-oriented' and maladaptive patterns as 'helpless' (Diener & Dweck, 1978, 1980; Dweck, 1975, 1986; Dweck & Leggett, 1988; Dweck & Reppucci, 1973; Weiner, 1985, 1986, 2018). Their goal was to uncover patterns that determined whether students avoided or approached challenges, how they responded to failure, and how long they persevered.

A tributary of the socio-cognitive approach posited that at the core of different motivational patterns were students' views about intelligence or ability (Dweck, 1986). Although cognitive ability exerts an enormous influence on academic performance (Deary et al., 2007), researchers and educators have long acknowledged that it does not account for all or most of the variance in academic performance (Duckworth et al., 2011, 2019; Thorndike, 1900). Factors (often identified by the misnomer 'noncognitive') like personality, test anxiety, motivation, and expectancies also play an essential role in scholastic achievement (Ackerman et al., 2013; Duckworth et al., 2019; McClelland, 1985). This realization formed the basis of the focus on "psychological factors, other than ability, that determine how effectively the individual acquires and uses [cognitive] skills" (Dweck, 1986, p. 1040). The theoretical shift that ensued was embedded within a broader framework of 'self-theories' (one’s views about the self); self-theories are purported to color the lens through which people see the world and affect the meaning of their experiences (Dweck, 2000). Eventually, Dweck and colleagues pinpointed a construct they hypothesized undergirds the different motivational patterns in academic contexts: beliefs about the *malleability* of intelligence, i.e., implicit theories of intelligence (ITOI; Dweck, 2006).

The Role of Intelligence

A discussion on the malleability of intelligence necessitates a preceding conversation about what is meant by the term ‘intelligence.’ Remarkably, but unsurprisingly, consensus about what intelligence *is* has yet to be reached. A coherent answer to the question, ‘What is intelligence?’ has evaded psychologists and philosophers alike. Theorizing on the nature of intelligence is rich and dates to ancient philosophers like Confucius (551–479 BCE), who defined intelligence as a set of skills, including verbal fluency and the ability to think flexibly (Pang et al., 2017), and Plato (428–348 BCE), who considered intelligence to be a love of truth and learning (Princiotta & Goldstein, 2015). In contemporary psychology, the measurement of intelligence progressed almost independently from theories of intelligence (Mackintosh, 2011). For decades, it was viewed simply as a psychometric tool, a technology devoid of theoretical or explanatory speculation about the nature of intelligence. A typical outlook in this tradition has been: “Intelligence is what the [intelligence] test tests” (Boring, 1923, p. 35). With the advent of factor analysis in 1904 by Charles Spearman, the psychometric tradition relied purely on statistical correlations among measures of mental faculties like reaction time, mental rotation, working memory, and verbal comprehension. Theories about the nature of intelligence have proliferated since, but most contemporary intelligence tests (e.g., Raven’s Progressive Matrices and Wechsler Intelligence Scales) continue to measure abilities like verbal and numerical fluency, abstract reasoning, and working memory—abilities honed through education.

That intelligence is central to the educational enterprise is hardly controversial. Individual differences in intelligence are most salient in the academic domain (Ritchie, 2015), and one of the most cited definitions of intelligence describes it as, among other things, the ability to “learn quickly and learn from experience” (Gottfredson, 1997a, p. 13). Learning—the primary purpose of education—and intelligence are thought to overlap considerably, and those with higher intelligence benefit more from direct instruction (Gottfredson, 1997b). Hence, the amount of education one attains is often determined by one’s intelligence, and one’s intelligence is reciprocally determined by the amount of education one attains. Academic performance is often used as a proxy measure for cognitive ability, and protracted, high-quality education remains one of the only consistent means of increasing individual intelligence (Brinch & Galloway, 2012; Dawson-Tunik et al., 2005; Ritchie et al., 2013; Ritchie & Tucker-Drob, 2018).

Implicit Theories of Intelligence

All explicit theories flow, one way or another, from implicit theories, which are lay beliefs about the fundamental nature of a phenomenon (Sternberg, 1990); in other words, they are "constructions by people (whether psychologists or laypersons) that reside in the mind of the individuals" (Sternberg, 1985, p. 608), with or without awareness of their existence. Implicit theories of intelligence have been studied extensively in relation to student motivation, and beliefs about the fundamental nature of intelligence are hypothesized to as meaning-making frameworks that result in different goals, behaviors, and attributions, leading to different affective, cognitive, and behavioral patterns (Dweck, 2006; Heyman & Dweck, 1992; Hong et al., 1999). A fixed view of intelligence has been linked to performance goals ('looking smart') and a tendency to explain failure in terms of low ability (Blackwell et al., 2007). Conversely, viewing intelligence as modifiable has been linked to mastery goals ('achieving mastery') and persistence in response to challenges and failures. Entity theorists (those who believe cognitive abilities are fixed) are purported to weigh ability more heavily for success, whereas incremental theorists (those who believe cognitive abilities are malleable) may be more likely to emphasize hard work (Hong et al., 1999). A sizable body of experimental and observational evidence suggests that when students believe intelligence to be a malleable entity, they select more challenging tasks, are more concerned about learning (vs. grades), respond better to failure, persist longer, and exhibit higher academic performance (Blackwell et al., 2007; Claro et al., 2016; McCutchen et al., 2016; Moser et al., 2011; Paunesku et al., 2015; Yeager et al., 2019).

Theoretical Background

The historical conceptualization of implicit theories of intelligence is rooted firmly in attribution theory; popular since the 1950s, attribution theory was first applied to the academic context by Bernard Weiner (Weiner, 1979; Weiner, 2018; Weiner & Kulka, 1970; see also Graham, 1991; Jones & Davis, 1965; Kelley, 1967; Thibaut & Riecken, 1955). Attributions are causal ascriptions of outcomes, and according to Weiner's theory, students attribute their academic performance to controllable or uncontrollable factors, with important implications for motivation. A related concept is *locus of control* (or internal-external control of reinforcement; Lefcourt, 1966; Rotter, 1954, 1975), which refers to whether individuals ascribe outcomes to internal causes (e.g., ability, effort) or external factors (e.g., luck, task difficulty). It is a generalized expectancy about whether one can influence one's outcomes, or in other words, the amount of personal responsibility shouldered for one's outcomes. Whereas the theory of locus of control focuses primarily on whether the attributions are internal or external, attribution theory addresses additional dimensions of controllability vs. uncontrollability and stability vs. instability. Research suggests that attributing outcomes to internal, controllable, and unstable causes is more likely to lead to adaptive motivational and emotional patterns (Weiner, 2018).

The second cornerstone in the foundation of implicit theories is the theory of learned helplessness. First developed using animal models, this theory was eventually applied to humans by Martin Seligman and colleagues (1972). Learned helplessness is a distortion of cognition, affect, and motivation that results from perceiving outcomes to be divorced from personal control. When individuals believe their actions to have no causal connection to outcomes, they manifest helpless behavior patterns. Carol Dweck first studied learned helplessness in children and found that holding ability constant, some children responded to failure with a helpless response (increased rate of error, lower persistence, and avoidance), whereas others remained resilient in the face of challenges and oriented towards achieving mastery (Dweck, 1975, 2011). Investigations suggested that children's mastery vs. helpless responses might be based on students' attributions for success and failure (Dweck, 2011). Students who attributed academic outcomes to internal and controllable factors (such as effort) were more likely to display mastery-oriented patterns (persevere and try harder), whereas those who attributed outcomes to external, uncontrollable causes were more likely to display helpless patterns and languish (Henderson & Dweck, 1990; Robins & Pals, 2002).

Eventually, Dweck and colleagues realized that conceptualization of the *nature of intellectual ability* influenced students' attributions for academic outcomes. Some considered intelligence a malleable trait that can be improved with effort, whereas others thought of it as an innate and unchanging entity. The terms *incremental* and *entity* theories of intelligence were thus coined, eventually supplanted by the descriptive terms 'growth mindset' and 'fixed mindset.' Although 'fixed' and 'growth' might imply reference to the stability vs. instability of intelligence—indeed, many refer to a fixed mindset as the belief that intelligence is stable (Sisk et al., 2018; T. Wilson et al., 2002)—researchers has explicitly stated that these beliefs are about *controllability* and not about the stability of intelligence (Dweck, 1975, 2011). Said differently, they are beliefs about whether one has the capacity to alter one's intelligence and not whether intelligence itself is susceptible to change (during development or as a function of age, for example). This perspective is aligned with a socio-cognitive approach to motivation, which heavily emphasizes a sense of personal agency (Bandura, 1989).

Along with differences in attributional ascriptions, empirical evidence suggests that different theories about intelligence also correlate with different goal orientations, which refer to the goals that motivate students in a classroom. Two are pertinent to implicit theories of intelligence: learning/mastery goals and performance/ego goals. (Note that in academic contexts, students also have relational and other non-academic goals.) When students are oriented toward learning, their engagement in academic tasks stems from a desire to learn and achieve mastery (R. M. Ryan & Deci, 2000). In contrast, students with performance/ego goals are motivated by a desire to 'appear smart' or 'not appear dumb,' mirroring the approach-avoidance conflict formulation (Atkinson, 1964; N. E. Miller, 1944, 1951, 1959). Research shows that students who hold performance goals are more likely to pick unchallenging tasks (to demonstrate ability), avoid asking for help (to maintain perceptions of ability), and seek answers (vs. understanding) when they do ask for help (Ames & Archer, 1988; Butler & Neuman, 1995; Meece et al., 2006). On the other hand, students with mastery goals are more likely to persevere, ask for hints (to achieve independent mastery), and are less concerned about the outward appearance of competence (Dweck, 1988).

Contemporary Issues

After decades of foundational academic research, the concept of 'growth mindset' has been shared broadly in a popular book (Dweck, 2006), trickled down to teacher professional development and curriculum redesigns (Yousaf, 2023), and eventually adapted into school and college interventions that inculcate an incremental view of intelligence (Paunesku et al., 2015; Yeager et al., 2013, 2016, 2019). Researchers have claimed that given the right context, social-psychological interventions can have long-lasting effects (Walton & Yeager, 2020; Yeager & Walton, 2011). A malleable view of intelligence is encouraged in the American education system by many as it is believed that holding such a view is beneficial for skill acquisition (Boaler, 2013). An item related to growth mindset has even been added to an international student assessment, Programme for International Student Assessment (PISA), yielding around 600,000 data points that suggest a positive relationship between growth mindset and math and reading achievement (Gouëdard, 2021; Sun et al., 2021).

However, the theory and its proliferation in the education system have not been without criticism. Multiple meta-analyses have been conducted, and the results are equivocal. Some indicate that light-touch growth mindset interventions might be less effective than proponents claim (Costa & Faria, 2018; Li & Bates, 2019; Macnamara & Burgoyne, 2023; Sisk et al., 2018). Others, using different analytical techniques, have reached different conclusions and issued calls for heterogeneity-sensitive methods (Burnette et al., 2023; Tipton et al., 2023). Although there are numerous valid criticisms about overreaching claims (Burgoyne et al., 2020) and the jury on social-psychological interventions is still out, the primary effect appears robust, despite small (but potentially meaningful; Funder & Ozer, 2019) effect sizes, especially for struggling students (Sisk et al., 2018).

Cross-Cultural Differences

Most research on implicit theories of intelligence has been conducted on Western samples. Although the concept has been highly influential in the American educational context, especially in the last decade, its generalizability to other cultures remains to be properly addressed. Pertinent to this discussion is intelligence research that addresses differences between conceptualizations of intelligence in different cultures (Berry, 1984; Pang et al., 2017; Yang & Sternberg, 1997). Scholars have long noted that cultural influences shape how intelligence is viewed since different behaviors count as intelligent in different sociocultural contexts (Cocodia, 2014; Gardner, 1983; Sternberg, 1985). Thus, the hallmarks of intelligence in non-Western cultures might be systematically different than in Western cultures, with a stronger emphasis on "social responsibility, hard work and perseverance" (Keats, 1982, p. 73).

In Eastern traditions, specifically Hindu and Buddhist philosophies that dominate many Eastern cultures, the concept of intelligence is inextricably linked to religious and moral domains (Das, 1994), and factors like determination, mental effort, and social intelligence are considered important manifestations of intelligent behavior (Gill & Keats, 1980; Okagaki & Sternberg, 1993; Super, 1983). High levels of achievement among Asian students have also been ascribed to a greater emphasis on educational attainment and reverence for a scholarly disposition (Cocodia, 2014; Stevenson & Stigler, 1992). Crucial differences have been identified between Eastern and Western perceptions about the importance of effort vs. innate ability in academic success. Chinese and Japanese students have been found to be more likely to attribute academic achievement to effort than American students, who are more likely to attribute it to talent or innate ability (Stevenson et al., 1993). Given the academic achievements of Asian and Asian-American students and the cultural importance placed by these communities on educational attainment, understanding cross-cultural differences in the conceptualization and perceived importance of intelligence in academic pursuits will be essential for developing motivation frameworks in this domain that can be generalized to other cultures.

Perceived Implicit Theories of Intelligence

According to expectancy-value theory, one of the most generative theoretical frameworks in motivation research, academic motivation depends on two factors, (a) whether students believe they can succeed (*expectancies*), and (b) whether they *value* the task/domain in question (Eccles & Wigfield, 1995; Tolman, 1938; Wentzel & Wigfield, 1998). Both are essential for engagement in behaviors that lead to academic achievement. If students do not believe that success is possible (that is, if they do not believe that their behavior is causally related to success), they are unlikely to expend effort, even if the domain is highly valued (a form of selfpreservation). Similarly, even if students believe that expending effort would lead to success, achievement-related behaviors are unlikely if students do not value achievement in that domain. Although personal expectancies (also called outcomes expectancies) are extremely important, students are also influenced by *others'* expectations regarding their ability.

Attribution theory, one of the forerunners of implicit theories, focused originally on "perceived causes of other persons' behavior" (Kelley & Michela, 1980, p. 458). Atkinson (1964) has noted that achievement motivation is at play "only when an individual knows that his performance will be evaluated (by himself or by others) in terms of some standards of excellence" (p. 240). In evaluative contexts, self-perceptions of ability matter, but so do perceptions of others' evaluation of one's ability or potential ability. As Bandura (2001) states, "human functioning is rooted in social systems. Therefore, personal agency operates within a broad network of socio-structural influences." (p. 14). In education—arguably the most consequential evaluative context of all—educator beliefs and expectations can manifest in differentiated behavior toward students in terms of the quality of interactions, provision of feedback, and the way student behavior is managed (Rubie-Davies, 2007). Teacher behavior, in turn, influences students' self-perceived ability, motivation, level of effort, and, ultimately, performance (T. L. Good, 1987; Graham, 1991; Rosenthal, 1985, 1987, 1991; Rosenthal & Jacobson, 1968a, 1968b).

Research on interpersonal expectancy effects indicates that teacher expectations of student performance may turn into self-fulfilling prophecies, resulting from mechanisms that include quality of feedback, response to failure, and individualized instruction (Rosenthal & Jacobson, 1966; Rosenthal & Rubie-Davies, 2015; Rosenthal & Rubin, 1978). Although research in the implicit theories domain has traditionally focused on individuals' personal implicit theories, recent advances have begun to address individuals' perceptions of their instructors' mindsets and the downstream psychological and behavioral consequences on the perceiver (Canning et al., 2019; Murphy & Dweck, 2010; Rattan et al., 2012).

Instructors' implicit theories may bleed into the classrooms through how they respond to struggles and mistakes and whether they attribute success in their classrooms to ability or to hard work. Entity theorists (those who believe cognitive abilities are fixed) may weigh ability more heavily for success. In contrast, incremental theorists are more likely to emphasize hard work (Hong et al., 1999). Teachers with incremental theories are more likely to emphasize that mistakes are learning opportunities (Walton & Yeager, 2020; Yeager et al., 2022). Regardless of teachers' implicit theories and how they manifest, the link between what educators think and what students *believe* educators think may not be exact. For example, an instructor might hold the implicit belief that students are capable of significant intellectual growth, but it may not reflect in their behavior or be perceived that way. For example, unsolicited help can inadvertently function as a cue of low ability (Graham & Chen, 2020), and classroom interactions can send subtle signals to students about their ability and potential outside the teacher's awareness. Such perceptions might shape how students view their academic potential, and students' expectancies affect academic behavior, mediated through influence on motivational systems (Henderson & Dweck, 1990; Masten & Coatsworth, 1998). Thus, it is crucial to measure *perceived* implicit theories of intelligence.

Perceived Implicit Theories of Intelligence (P-TOI) refers to students' perceptions of their instructors' beliefs about the malleability of intelligence. The construct has been postulated recently under various labels: *meta-lay theories about intellectual potential* (Rattan et al., 2018), *perceived faculty growth mindset* (Muenks et al., 2020), and *students' perceptions of instructors' mindset beliefs* (Kroeper et al., 2022). Preliminary evidence suggests that these perceptions or meta-lay theories correlate with important psychological and academic outcomes (engagement, persistence, concerns about negative evaluation, and performance). More specifically, findings indicate that perceiving an instructor to hold a fixed view of ability correlates with greater psychological vulnerability—lower sense of belonging, higher imposter feelings, higher evaluative concern—and lower engagement and performance in STEM (Science, Technology, Engineering, and Mathematics) courses (Canning et al., 2019; Kroeper et al., 2022; Muenks et al., 2020, 2021). Other work has shown that instructors who hold an entity view of intelligence are more likely to comfort students for poor performance, leading to lower motivation, and to consider a single poor performance a sign of low ability/potential (Rattan et al., 2012). Together with the finding that students may not be able to *accurately* assess teachers' theories about intelligence (Haimovitz & Dweck, 2017), the importance of ascertaining which cues from educators indicate an entity vs. incremental view of intelligence is heightened (Kroeper et al., 2022).

Given this construct's interpersonal, dyadic nature, I hypothesize that an adaptive behavior especially likely to be influenced by perceived implicit theories of intelligence is *academic help-seeking* (e.g., asking questions, visiting office hours, and seeking tutoring). As I delineate below, the social nature of help-seeking, the threats to independent mastery that may be associated with it, and its central role in the conceptualization of implicit theories of intelligence make it a crucial but understudied construct in the implicit theories nomological network.

Academic Help-Seeking

Academic help-seeking is a crucial self-regulatory strategy linked to a host of positive outcomes, including improvement in metacognitive skills, academic engagement, and achievement (Butler, 1998; Karabenick & Knapp, 1988, 1991; Karabenick & Sharma, 1994; Nelson-Le Gall & Glor-Scheib, 1985; Newman, 1990, 1994). According to Newman (2002), several competencies and motivational factors are crucial for adaptive help-seeking—cognitive competencies (knowing when help is necessary, what questions to ask), social competencies (knowing whom to ask and how to do it in socially appropriate ways), contextual motivational resources (e.g., student-teacher interactions and grading systems), and personal motivational resources (e.g., personal agency and willingness to express a need for help). Help-seeking behavior, thus, depends on both environmental and personal variables, including personal and classroom goals, task difficulty, self-beliefs, collaborative activities, and cultural context (Deci & Ryan, 1987; Markus & Kitayama, 1991; A. M. Ryan et al., 1998; Tessler & Schwartz, 1972).

Of interest in relation to theories of intelligence are reasons why students might not avail or accept help, the "help-seeking dilemma" (Nadler, 1997, p. 379). Despite increased metacognitive abilities, students can be reluctant to seek help even when they know they need it, especially during adolescence (A. M. Ryan et al., 2001). Like most individuals, students can be reluctant to ask for help when asking for help is psychologically costly. According to Nadler (1997), students are hesitant to seek help when the cost of requesting help outweighs the benefits of overcoming difficulty. Several theoretical parallels exist between maladaptive ("helpless") motivational patterns and help-seeking avoidance. Within a socio-cognitive framework, the focus is primarily on *psychosocial* factors, and through a series of ingenious experiments, Butler (1998) revealed two primary orientations that can result in help-seeking avoidance: a desire for autonomy and a threat to competence. Asking for help can conflict with a need for autonomy and violate the “norm of self-reliance, which is so well ingrained in Western civilization” (Nadler & Fisher, 1986, p. 82; see also McClelland et al., 1953). Indeed, help-seeking was initially viewed as a sign of dependency even within the research literature (Butler, 2006). Hence, if help-seeking is perceived as a sign of incompetence, especially in one’s cultural context, it can be uncomfortably threatening to self-perceived ability.

A.M. Ryan and Pintrich (1997) found that many students worried about negative judgments from their teachers and classmates regarding their abilities. Lower perceptions of perceived cognitive competence, but also low perceived social competence, could result in avoidance of help-seeking. Thus, students can be reluctant to seek help if it threatens their self-esteem (Fisher et al., 1982; Nadler & Fisher, 1986). This reluctance might be especially relevant for perceived theories of intelligence since perceiving an instructor to believe that “some people are smart, others are not” has the potential to compound this threat. As Jussim (1990) highlights, students with lower confidence in their abilities might be more susceptible to this effect:

> (S)tudents who know they are smart may have the resources for overcoming a teacher’s initially erroneous belief that they are dumb. Students who are just as objectively intelligent, however, but who have less clear self-conceptions of their academic skills, may be more subject to expectancy effects. (p. 24).

Regarding help-seeking specifically, A.M. Ryan et al. (2001) summarize:

> Students who feel insecure about their abilities—academically or relating socially to other students—are more likely to avoid help seeking. Students who are focused on their reputation—academic or social—are more likely to avoid help seeking. In classrooms where teachers emphasize personal improvement and promote positive social relationships, concerns about help seeking decrease and concerns about help avoidance decrease. In contrast, in classrooms where teachers highlight ability comparisons among students, concerns about help seeking increase and ... help avoidance increase(s). (p. 111)

Butler (1998) hypothesized that different orientations toward help avoidance would determine different kinds of help-seeking behavior. To test this hypothesis, she asked 1,029 10 to 12-year-olds to rate their reason for avoiding asking for help in math and categorized the responses into three factors: autonomous striving for independent mastery (“I want to do it on my own”); ability focuses concerns to mask poor ability (“I don’t want to look stupid”); and expedient perceptions that help would not expedite task completion (“I wouldn’t just get the ‘answers’”). Butler further predicted that the three orientations would lead to different responses to academic difficulty, which was confirmed by a second study. Students who had an autonomous orientation asked for help when they could not solve the problem, asked for hints (rather than explicit directions to get to the answer), and spent more time working on the problem alone. Those with an expedient orientation to help-seeking avoidance spent less time working on problems and were more likely to ask for directions. Those with ability-focused concerns were more likely to use an avoidant-covert style (copying correct answers instead of asking for help).

Attribution theory predicts that students would be less likely to seek help when the need for help is attributed to internal causes (e.g., low ability or competence) rather than external (difficult task; Tessler & Schwartz, 1972). Ames and Lau (1982) looked at the decision to seek help as a function of past performance, attribution of past performance (help-relevant vs. help-irrelevant attribution), and availability of information regarding whether seeking help would be beneficial. Administering a behavioral measure for help-seeking—attendance of a help session before an exam—to college students, they found that those more likely to go to help sessions were those who had done poorly in the past and attributed the poor performance to factors within their control. Students' goal orientations (mastery vs. performance) have also been linked to help-seeking behavior such that mastery-orientated students have fewer concerns about perceived competence and, therefore, lower help-seeking avoidance. On the other hand, performance-goal orientation increases competence concerns and, consequently, increases help-seeking avoidance. Given that an incremental mindset is associated with mastery orientation and an entity mindset with performance orientation (Dweck, 2000; Licht & Dweck, 1984), it is reasonable to expect a fixed view of intelligence to correlate with help-seeking avoidance.

In terms of environmental determinants, classroom norms have been linked to academic help-seeking—students feel comfortable seeking help when needed in contexts where seeking and providing help is valued and incentivized. On the other hand, environmental constraints outside an individual's control can hinder help-seeking; for example, if there is no willing or able helper, it will take too long, or the setting is not right. Micari and Calkins (2021) examined the relationship between instructor openness and student questions, help-seeking behavior, and final grades in courses. They found that perceived instructor openness and help-seeking were positively related to course grades. A.M. Ryan and colleagues (2001) evaluated individual and classroom factors—classroom goal structures, teachers' support of students' social-emotional needs, and students' self-efficacy—and how they related to help-seeking avoidance. They found that students' self-efficacy was negatively correlated with help-seeking avoidance, but this relationship was less salient in classrooms where teachers cared about students' social and emotional needs. Classrooms that focused more on understanding and mastery (as opposed to competition) and less on proving one's ability also exhibited lower levels of help-seeking avoidance.

In contexts where help is readily available, students' reluctance to seek help can be particularly troubling for educators and leave them scratching their heads (Grayson et al., 1998; Tessler & Schwartz, 1972). Bohns and Flynn (2010) have suggested that helpers overestimate the likelihood that people will ask for help and underestimate the role of embarrassment in help-seeking avoidance. Individuals' perceptions of their environment and themselves also have important implications for whether they seek help (Grayson et al., 1998). Therefore, looking at help-seeking behavior from a dual lens of perceptions of one's attributes and attributes of the context is in order.

Help-Seeking and Implicit Theories of Intelligence

Recent critiques of implicit theories of intelligence claim that the theory suggests that students simply need to "put in more effort." In response, researchers have argued that having a growth mindset is more than just hard work—students must try different strategies and ask for help when stuck (Dweck, 2016). In fact, seeking help from others is often baked into the description of what it means to hold an incremental view of intelligence:

> "Individuals who believe their talents can be developed (through hard work, good strategies, and *input from others* [emphasis added]) have a growth mindset." (Dweck, 2016, p. 2)

> "…people who hold more of a growth mindset endorse the belief that intelligence is malleable and can be expanded and developed by persistence, *help-seeking* [emphasis added], and adopting the right strategies." (Muenks et al., 2020, p. 2)

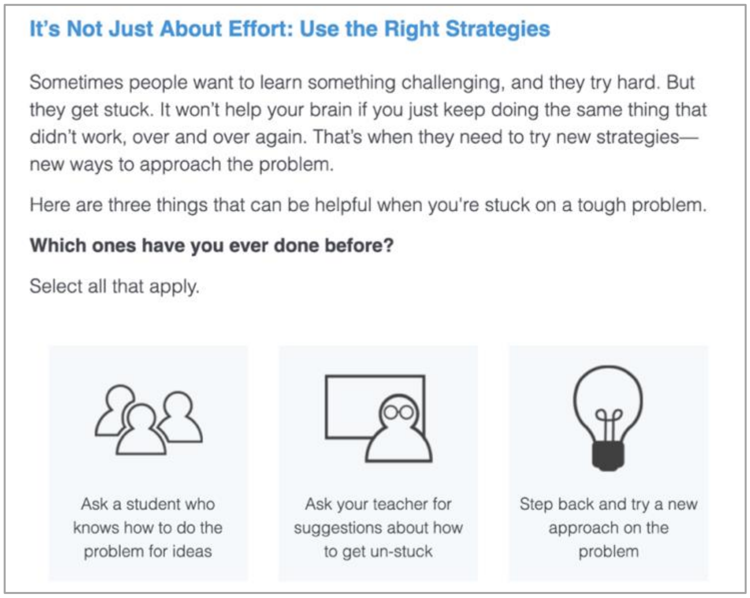
> "For instance, a growth-mindset-of-intelligence intervention conveys to students the malleability of intellectual abilities in response to hard work, effective strategies, and *help from other people* [emphasis added]." (Yeager et al., 2022, p. 18)

A similar pattern can be observed in widely-used interventions that encourage students to adopt an incremental view of intelligence (Dweck & Yeager, 2019; Yeager et al. 2019). Along with sharing empirical evidence about brain plasticity, these interventions highlight the importance of trying different strategies and not just expending effort. Help-seeking is recommended as one of the most salient active strategies as part of these interventions: "If you are stuck on a problem, *ask a student* who knows how to do the problem for ideas, or *ask your teacher* [emphasis added] for suggestions on how to get unstuck" (See Figure 1). Although little empirical work on the relationship between growth mindset and help-seeking behavior exists, past research on academic help-seeking suggests that those who are most likely to hold an entity belief about intelligence might also be less likely to ask for help due to differences in goal orientation: mastery ("I want to learn") vs. performance ("I want to look competent"). Consequently, those most likely to benefit from seeking help from others might be least likely to ask for it.

Despite its importance in defining *and* encouraging a malleable view of intelligence, there is virtually no empirical research looking directly at the relationship between implicit theories of intelligence and academic help-seeking. The question of how personal and perceived implicit theories about intelligence relate to help-seeking behavior represents a theoretical and empirical gap that the current research hopes to help address.

Figure 1

Screenshot from a Growth Mindset Intervention



Note. Screenshot from Yeager et al. (2019).

CHAPTER II

FACULTY GROWTH MINDSET AND ACADEMIC HELP-SEEKING

Structured Abstract

Background: Although help-seeking is theorized as a crucial adaptive behavior tied to a malleable view of intelligence, little empirical work has directly addressed the connection between academic help-seeking and implicit theories of intelligence.

Purpose: To test whether students' perceptions of their instructors' theories about intelligence correlate with students' self-reported behavior and attitude toward help-seeking.

Participants: 1,662 college students recruited from two North American Universities ($n_{UniversityA}$ = 785; $n_{UniversityB}$ = 365) and an online data collection platform, Prolific (https://www.prolific.co; *n* = 512).

Research Design: Cross-sectional survey

Data Collection and Analysis: Data were collected between March–December 2020 via Qualtrics (https://www.qualtrics.com). Analysis was conducted in R (R Core Team, 2023), ACER Conquest (Adams et al., 2012), and BEAR Assessment System Software (BASS; Wilson & Sloane, 2000). Hypotheses and analysis plan were not preregistered.

Findings: Students who perceived instructors at their institutions to hold a malleable view of intelligence were more likely to report positive attitudes towards help-seeking, controlling for a host of psychological and demographic factors. These perceptions robustly predicted many of the psychological variables measured in the study. Academic help-seeking is best predicted by students' level of academic difficulty and sense of self-efficacy.

Conclusion: In this study, we found that perceptions of instructors' theories about intelligence indeed correlate with students' attitudes toward help-seeking but not with self-reported help-seeking behavior.

Faculty Growth Mindset and Academic Help-Seeking

The current study aimed to test whether there exists an empirical link between students' perception of their instructors' implicit theories about intelligence and their attitude toward help-seeking. The term *attitude* is employed in the technical sense, denoting the tripartite affect-behavior-cognition—the ABC—model of attitude structure (Breckler, 1984); we measured students' beliefs about, affective responses to, and self-reported engagement in academic help-seeking.

Of interest in the current investigation were several theoretically relevant psychological and demographic factors that could moderate the relationship, if it exists, between academic help-seeking and perceived theories of intelligence. At the top of this list of factors are students' personal theories about intelligence, of which perceived theories of intelligence might simply be a reflection. It is possible that students, especially college students, do not have sufficient information to make judgements of their instructors' theories of intelligence. In that case, students might impute their own views about intelligence in response to queries about their perceptions of their instructors' implicit theories. We measured students' personal theories of intelligence to test this hypothesis and included it as a covariate in all models.

Students' academic self-efficacy, which has been positively linked to help-seeking, might determine their facility in identifying the need for help and the confidence to successfully seek it (A. M. Ryan et al., 2001). Furthermore, whether students ask for help likely depends on the amount of help needed. There is evidence of a curvilinear (inverted U-shaped) relationship between help-seeking and academic need; those least or most in need of assistance may be least likely to seek it (Karabenick & Knapp, 1988; Rosen, 1983). Thus, we also hoped to account for the level of academic difficulty faced by the students in our sample when interpreting our main results.

We were partly interested in replicating findings from previous studies that have linked students' perceptions of their college instructors' theories about intelligence (Perceived Faculty Growth Mindset) with psychological constructs like belonging uncertainty, impostor phenomenon, negative affect, dropout intentions, and perceived competitiveness of the college environment (Canning et al., 2019; Muenks et al., 2020). In terms of demographic characteristics, we were interested primarily in first-generation college student status. According to the cultural mismatch theory (Stephens et al., 2012), the performance gap between first-generation and continuing-generation college students results from a mismatch between the interdependent cultural backgrounds of first-generation students and the independent norms of universities. Since asking others for help is a social activity that embodies interdependent cultural values, we wanted to explore whether there was a difference in first-generation students' approach to help-seeking.

We also expected a moderate positive correlation between students' own theories of intelligence (i.e., their growth mindset) and their perceived faculty growth mindset. In addition to replicating effects found in previous work, we made several novel predictions; if there is a relationship between faculty growth mindset and help-seeking, as we expected based on theoretical considerations, perceived faculty growth mindset should positively predict students' attitude toward help-seeking. Finally, we expected perceived competitiveness and higher impostor feelings to predict lower help-seeking behavior and for students higher in growth mindset to view help-seeking as less threatening.

Method

Participants

Participants were college undergraduates (M_{age} = 21.51, SD_{age} = 4.42; 62% female), recruited from two public North American universities ($n_{UniversityA}$ = 785; $n_{UniversityB}$ = 365) and an online survey platform, Prolific (https://www.prolific.co; n = 512, limited to American college students). The study was approved by the Institutional Review Boards (IRB) at both universities A and B. Participation took 27 minutes on average and was compensated with either course credit or US $3.50. Table 1 shows the demographic composition of the sample.

Since we relied primarily on research participant pools, we did not establish a stopping rule and recruited as many students as possible across one academic year (including the summer). Online recruitment was determined by the amount of funding available. Post-hoc power analysis for multiple regression (using the *pwr* package in R; Champely, 2020) indicated that our final sample size (N = 1,662) allows us to detect an effect size of r = .01 or greater with 80% power and an alpha of .05.

Table 1

Sociodemographic Characteristics of Participants in Study 1

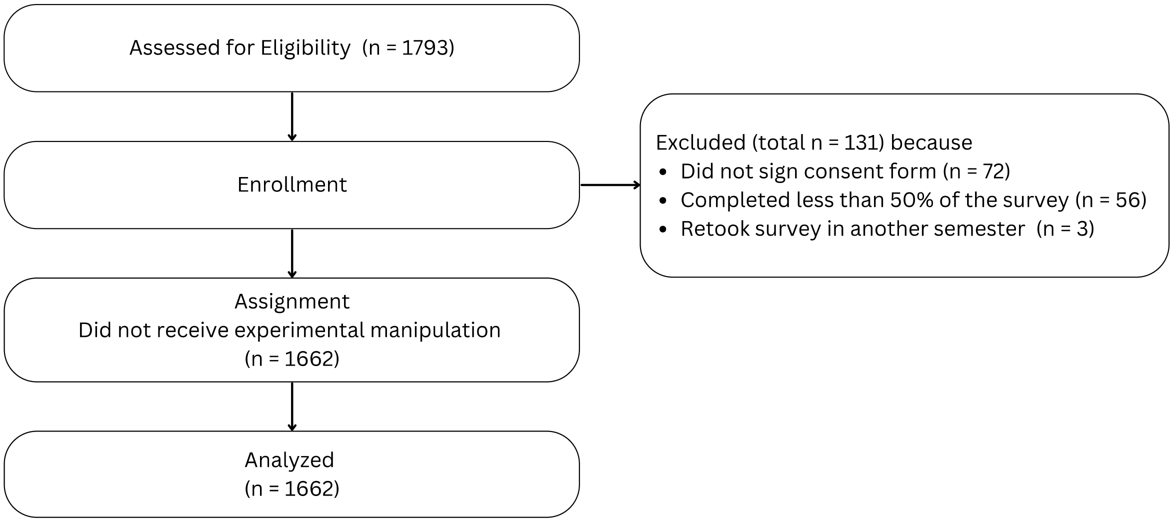
Baseline Characteristics	University A	University B	Prolific	Overall
Age	20.75 (3.01)	20.97 (3.41)	23.08 (6.16)	21.51 (4.42)
Gender				
Female	581 (74 .01%)	236 (64.66 %)	247 (48.24 %)	1064 (64.02 %)
Male	183 (23.31 %)	125 (34.25 %)	238 (46.48 %)	546 (32.85 %)
Non-binary/Other	15 (1.91 %)	3 (0.82 %)	15 (2.93 %)	33 (1.99 %)
Missing	6 (0.76 %)	1 (0.27 %)	12 (2.34 %)	19 (1.14 %)
Generation Status				
Non-First Generation	559 (71.21 %)	264 (72.33 %)	362 (70.70 %)	1185 (71.30 %)
First Generation	220 (28.03 %)	60 (16.44 %)	139 (27.15 %)	419 (25.21 %)
Missing	6 (0.76 %)	41 (11.23 %)	11 (2.15 %)	58 (3.49 %)
Race/Ethnicity				
East Asian	229 (29.17 %)	121 (33.15 %)	50 (9.77 %)	400 (24.07 %)
South Asian	67 (8.54 %)	53 (14.52 %)	24 (4.69 %)	144 (8.66 %)
Southeast Asian	83 (10.57 %)	32 (8.77 %)	31 (6.05 %)	146 (8.78 %)
Black/African/African-American	10 (1.27 %)	10 (2.74 %)	50 (9.77 %)	70 (4.21 %)
Hispanic/Latinx	97 (12.36 %)	8 (2.19 %)	60 (11.72 %)	165 (9.93%)
Middle Eastern	20 (2.55 %)	21 (5.75 %)	7 (1.37 %)	48 (2.89 %)
Native American or Alaskan	1 (0.13 %)	3 (0.82 %)	1 (0.20 %)	5 (0.30 %)
Pacific Islander	5 (0.64 %)	1 (0.27 %)	0 (0.00 %)	6 (0.36 %)
White/European	194 (24.71 %)	87 (23.84 %)	243 (47.46 %)	524 (31.53 %)
Biracial/Mixed	65 (8.28 %)	19 (5.21 %)	31 (6.05 %)	115 (6.92 %)
Not Listed	5 (0.64 %)	8 (2.19 %)	3 (0.59 %)	16 (0.96%)
Missing	9 (1.15 %)	2 (0.55 %)	12 (2.34 %)	23 (1.38 %)
Social Class				
Poor	23 (2.9%)	1 (0.3%)	18 (3.5%)	42 (2.5%)
Working Class	102 (13.0%)	19 (5.2%)	72 (14.1%)	193 (11.6%)
Middle Class	617 (78.6 %)	324 (88.8%)	402 (78.5%)	1343 (80.8%)
Upper Class	35 (4.5%)	18 (4.9%)	7 (1.4%)	60 (3.6%)
Missing	8 (1.0%)	3 (0.3%)	13 (2.5%)	24 (1.4%)

Inclusion and Exclusion

Participants who signed the consent form and completed at least half of the survey were included in the study. For the online sample, we aimed to recruit 498 participants and received 540 responses before the survey closed, out of which 512 responses had at least a 50% completion rate and were included in the analysis. (Note that 50% was an arbitrary choice; reducing this percentage to 20 increases the sample size to 1673 but provides less complete data for the instruments collected.) Three participants from University A completed the survey during both data collection waves, and their first observations were retained. No other observations were excluded (see Figure 2 for participant flowchart).

Figure 2

CONSORT Flowchart of Participants in Study 1



Three embedded attention checks asked participants to pick a particular response option (e.g., "Please select '*Slightly characteristic of me*'"). The three attention checks were failed by 7%, 14%, and 40% of the participants, respectively. The attention check with the lowest pass rate (60%) was a standalone item with a longer, more involved stem, and only participants paying close attention would have noticed the instruction to pick a specific response. 81% of the participants passed the other two attention checks, and only 52% of the participants passed all three attention checks. We do not exclude any participants in the main analyses but report the results after limiting the sample to participants who passed two out of the three (n = 1353) as well as all three attention checks (n = 857) in Appendix B.

Procedure

Participation occurred online through Qualtrics (https://www.qualtrics.com) between March 9–December 2, 2020. After consenting to the study, participants completed self-report measures, clustered by instrument with instrument order randomized, and the survey concluded with a demographic questionnaire.

Data for this study were collected in collaboration with another research team also interested in college students' psychological experience, allowing us to collect a larger sample. The overlap between our theoretical interests meant that most instruments/items administered by our collaborators are potentially related to our primary variables. Although not in line with the original plan, we decided to incorporate constructs measured by the other team in this study for several reasons: first, the purely exploratory nature of the study and the relevance to the constructs under question for our research questions renders a distinction between the two sets of measures arbitrary and a claim of irrelevance to the current study untrue. Second, it allows us to make complete use of available data and get a fuller picture of students' psychological experiences. Finally, it allows us to be completely transparent and report everything that was collected. Thus, we see little drawback in including all collected measures in this report.

Instruments

Table 2 provides a descriptive list of all instruments and items (including demographic items) included in the survey, along with information regarding which items were included in the survey during each of the three waves (Spring, Summer, and Fall 2020). Here, we provide additional information about the primary constructs of interest.

Faculty Growth Mindset

Faculty Growth Mindset refers to students' perceptions of their instructors' theories about intelligence and is the primary predictor (independent variable) in the current study. It was measured using four out of five items developed by Muenks and colleagues (2020), adapted from the original growth mindset items (Dweck, 2000) by changing the item wording from asking about one's beliefs about intelligence ("You have a certain amount of intelligence, and you really can't do much to change it.") to perceptions of the instructor's beliefs about intelligence ("The professor in this class seems to believe that students have a certain amount of intelligence, and they really can't do much to change it"). Muenks et al. administered these items to students in a particular course, whereas our study asked students about their perceptions of the instructors at their institution at large. Thus, we adapted the items slightly to reflect this change in focus (e.g., "In general, most professors at my institution seem to believe that students have a certain amount of intelligence, and they really can't do much to change it.")

Help-Seeking

Help-Seeking (HS) was measured in several ways, relating to the ABC model of attitude structure. We administered seven items, created ad hoc, pertaining to instances Help-Seeking Behavior in the previous month (e.g., going to office hours, visiting the tutoring center, and asking other students for help). Given that the start of the COVID-19 pandemic and our data collection coincided, we discontinued administering these items after the first wave of data collection and have data from only 626 participants (university A and B from Spring 2020). Self-reported help-seeking behavior was alternatively assessed using a 5-item subscale from the Motivated Self-Regulated Learning Questionnaire (MSLQ; McKeachie et al., 1985; Pintrich & DeGroot, 1990) that inquired about students' general tendency to seek support from peers and instructors. Two 3-item instruments from Skaalvik & Skaalvik (2005) were used to assess whether students considered help-seeking to be threatening (Help-Seeking as Threatening) and how they perceived their own help-seeking behavior (Self-Perception of Help-Seeking Behavior). Three other items were created ad hoc for this study, assessing students' general attitude toward help-seeking (Perception of Help-Seeking), two of which are included in a composite index.

Thus, the Help-Seeking Composite includes information from 13 items—MSLQ (5), Help-Seeking as Threatening (3), Self-Perceptions of Help-Seeking Behavior (3), and Perceptions of Help-Seeking (2). A 14th item was also created for this study (Self-Help) but is not included in the composite as it was not administered in the first data collection wave. The composite was created using an item-response partial credit model (Masters, 1982), and the weighted likelihood estimators (EAP person estimates) were used as a measure of subjects' attitude towards help-seeking. (Please refer to Chapter 3 for a description of the model and person estimates.)

Academic Self-Efficacy

Academic Self-Efficacy refers to students' perceived capability to successfully carry out academic tasks and was assessed using 11 items that measured students' use of adaptive self-regulated learning strategies (Bandura, 1989; Zimmerman et al., 1992). The response scale for the instrument was adapted from a 7-point Likert scale to a 100-point 'confidence' scale ("How confident are you that you can...") based on the recommendation by Bandura (2006) for constructing self-efficacy instruments. The response range was presented from 0 to 100 in 10-point increments, yielding an 11-point scale.

Impostor Phenomenon

Impostor Phenomenon (IP) refers to feelings of personal incompetence and fraudulence despite high achievement (Clance, 1985; Clance & Imes, 1978). It was measured using the Clance Impostor Phenomenon Scale (CIPS; Clance, 1985), a 20-item instrument with a 5-point Likert scale that measures three facets of impostor phenomenon: feeling like a fake, discounting achievement, and attributing success to luck.

Analysis Plan

We approached data analysis in three ways. Our primary approach was to use multiple regressions and bivariate correlations for testing the main hypotheses. Given the large number of covariates collected in the study, the two secondary approaches homed in on the main independent variable, Faculty Growth Mindset, and the main dependent variable, attitude toward Academic Help-Seeking. To test the explanatory power of the primary independent variable, Faculty Growth Mindset, we conducted a specification curve analysis and simultaneously ran regression models that included all the psychological variables as dependent variables and all the covariates as controls. Lastly, to assess the best predictors of Academic Help-Seeking, we used regularized regressions to isolate the covariates that explained most of the variance in attitude toward Academic Help-Seeking.

Open Research Practice Statement

The study hypotheses and analysis plan were not preregistered. The entire survey, anonymized data, and code are publicly available at https://researchbox.org/870.

Table 2

Instruments/Items Included in Data Collection Waves for Study 1

Instrument	Cronbach's α	Description	Uni A (W1)	Uni B (W1)	Prolific (W2)	Uni A (W3)
Help-Seeking	.72	5-item subscale from the Learning Strategies section of The Motivated Strategies for Learning Questionnaire (MSLQ) that measure students' tendency to manage support and seek help with coursework. *Source*: Pintrich & DeGroot (1990) *Response options*: 7-point Likert scale from 1 = *not at all true of me* to 7 = *very true of me* *Example item*: "I ask the instructor to clarify concepts I don't understand well."	X	X	X	X
Help-Seeking as Threatening	.87	3 items that measure students' concerns about being assessed negatively by instructors and peers for seeking help *Source*: Skaalvik & Skaalvik (2005) *Response options*: 5-point Likert scale from 1 = *false* to 5 = *true* *Example item*: "I worry that other students may think that I am stupid if I ask for help."	X	X	X	X
Perceptions of Help-seeking	.48	2 items that assessed students' general evaluation of academic help-seeking *Source*: Ad hoc *Response options*: 6-point Likert scale from 1 = *strongly agree* to 6 = *strongly disagree* *Example item*: "If a student visits the tutoring or academic help center a lot, that means they probably won't do well in their classes."	X	X	X	X
Self-perceptions of Help-seeking Behavior	.80	3 items measuring students' perceptions about their own help-seeking behavior *Source*: Skaalvik & Skaalvik (2005) *Response options*: 5-point Likert scale from 1 = *false* to 5 = *true* *Example item*: "I do not ask for help even when I need it."	X	X	X	X
Help-seeking Behavior	.75	7 items that asked student to report how often they sought help from instructors, peers, and institutional resources in the previous month *Source*: Ad hoc *Response options*: 4-point Likert scale from 1 = *1-2 times* to 4 = *5 times or more* *Example item*: "In the past month, how often did you: go to a professor's office hour?"	X	X		
Self-help		"When stuck, I usually try to figure things out on my own before asking the professor or teaching assistant for help." *Source*: Ad hoc *Response options*: 6-point Likert scale from 1 = *strongly agree* to 6 = *strongly disagree*			X	X
Perceived Faculty Growth	.82	4-item instrument that measures students' perceptions of their instructors' beliefs about the malleability of intelligence	X	X	X	X

Instrument	Cronbach's α	Description	Uni A (W1)	Uni B (W1)	Prolific (W2)	Uni A (W3)
Mindset†		*Source*: Muenks et al. (2020) *Response options*: 6-point Likert scale from 1 = *strongly agree* to 6 = *strongly disagree* *Example item*: "In general, most professors at my institution seem to believe that students have a certain amount of intelligence, and they really can't do much to change it."				
Growth Mindset†	.91	3-item instrument that assesses beliefs about the malleability of intelligence *Source*: Dweck (2000); Yeager et al. (2016) *Response options*: 6-point Likert scale from 1 = *strongly agree* to 6 = *strongly disagree* *Example item*: "Your intelligence is something about you that you can't change very much."	×	×	×	×
Academic Self-Efficacy†	.90	11-item instrument measuring students' level of confidence about successfully accomplishing academic tasks *Source*: Zimmerman et al. (1992) *Response options*: 11-point slider scale from 0 to 100 (10-point intervals) from 0 = *no confidence at all* to 100 = *complete confidence* *Example item*: "How much confidence do you have that you can successfully: finish assignments by deadlines?"			×	×
Impostor Phenomenon†	.91	20-item instrument, Clance Impostor Phenomenon Scale (CIPS), that assesses fear of evaluation and failure, feelings of inadequacy, inability to internalize success, and worry that others will discover one's lack of ability. *Source*: Clance (1985) *Response options*: 5-point Likert scale from 1 = *not true at all* to 5 = *very true* *Example item*: "I can give the impression that I'm more competent than I really am."	×	×	×	×
Belonging	–	"I belong at [institution]." *Source*: Walton & Cohen (2007) *Response options*: 6-point Likert scale from 1 = *strongly disagree* to 6 = *strongly agree*	×	×	×	×
Belonging Uncertainty	.82	3 items measuring students' level of uncertainty about belonging in college *Source*: Broda et al (2018); Yeager et al. (2016) *Response options*: 5-point Likert scale from 1 = *not at all true of me* to 5 = *completely true of me* *Example item*: "Sometimes I worry that I do not belong in college."	×	×	×	×
Self-Determination Theory†*	.90	Basic Needs Satisfaction in General Short Scale (BNSG-S) was used to measure the three basic needs postulated by Self-Determination Theory: autonomy, competence, and relatedness *Source*: Deci & Ryan (2000); Gagné, (2003) *Response options*: 7-point Likert scale from 1 = *not at all true* to 7 = *very true* *Example items*: A = "I feel like I can pretty much be myself in my daily situations."; C = "Most days I feel a sense of accomplishment from what I do." R = "People in my life care about me."	×	×	×	×

Instrument	Cronbach's α	Description	Uni A (W1)	Uni B (W1)	Prolific (W2)	Uni A (W3)
Perceptions of Competitiveness	.68	Two items measuring perceived level of competitiveness at one's university *Source*: Canning et al. (2019) *Response options*: 7-point Likert scale from 1 = *strongly agree* to 7 = *strongly disagree* *Example item*: "The professors at this institution seem to pit students against each other in a competitive manner."	X	X	X	X
University Commitment*	.62	3 items assessing students' level of attachment to university *Source*: Unknown *Response options*: 6-point Likert scale 1 = *strongly agree* to 6 = *strongly disagree* *Example item*: "I do not feel "emotionally attached" to [my university]."	X	X	X	X
Dropout Intentions		"This semester, how often have you thought about dropping out of school?" (*Never / Rarely / Sometimes / Often / Always*)	X	X	X	X
Academic Difficulty		"In the past month, have you faced academic difficulty?" (*Yes/No*)	X	X	X	X
Sense of Self*	.87	12-item Sense of Self Scale (SOSS) measuring four components: lack of understanding of oneself; sudden shifts in feelings, opinions, and values; difficulty keeping one's identity separate from that of others; and feeling of a tenuous existence. *Source*: Flury & Ickes (2005) *Response options*: 6-point Likert scale from 1 = *very uncharacteristic of me* to 6 = *very characteristic of me* *Example item*: "I have a clear and definite sense of who I am and what I'm all about."	X	X	X	X
Goal Motivation*	.88	6-item instrument that first asks students to report an important goal (open-ended), followed by 5 items measure their level of motivation and commitment towards the goal *Source*: Unknown *Response options*: 6-point Likert scale from 1 = *strongly disagree* to 6 = *strongly agree* *Example item*: "I am motivated to pursue this goal."	X	X	X	X
State Authenticity as Fit to Environment†*	.93	15-item instrument that measures self-concept fit, goal fit, and social fit to the environment. *Source*: Based on Schmader & Sedikides (2018) *Response options*: 7-point Likert scale from 1 = *strongly disagree* to 7 = *strongly agree* *Example items*: SC = "[My university] feels right for who I am." G = "[My university] is a place where I feel intrinsically motivated by my own goals." S = "When I'm around other students on campus, I feel like I can act natural."	X	X	X	X
Collective Self-Esteem (CSE)*	.89	16-item instrument that measures students' relation to their social groups as it pertains to their experience in college; includes membership self-esteem, private collective self-esteem, public collective self-esteem, and importance to identity	X	X	X	X

Instrument	Cronbach's α	Description	Uni A (W1)	Uni B (W1)	Prolific (W2)	Uni A (W3)
		Source: Luhtanen, R., & Crocker, J. (1992) *Response options*: 7-point Likert scale from 1 = *strongly disagree* to 7 = *strongly agree* *Example items*: Membership = "I am a worthy member of the social groups I belong to on campus." Private = "I feel good about the social groups I belong to." Public = "In general, others on campus respect the social groups that I am a member of." Identity = "The social groups I belong to are an important reflection of who I am."				
Survey of Positive and Negative Experience (SPANE)*	+ (.90) – (.84)	12-item instrument measuring students' experience of positive and negative emotions/feelings *Source*: Diener et al. (2009) *Response options*: 5-point Likert scale from 1 = *very rarely* to 5 = *very often or always* *Item*: "Please think about what you have been doing and experiencing since coming to [institution]. Then report how much you experienced each of the following feelings: Positive, Negative, Good, Bad, Pleasant, Unpleasant, Happy, Sad, Afraid, Joyful, Angry, Contended"	X	X	X	X
State Authenticity*		"At [my institution], I feel…" Source: Ad hoc *Response options*: 7-point Likert scale from 1 = *inauthentic* to 7 = *authentic*	X	X	X	X
Real Self - Observed Self (RSOS)*		"Which pair of circles best represent how close you feel to your real self at [institution]?" Response option: Seven variations of overlapping circles denoting varying overlap between real and observed self (see Supplementary Material for details)	X	X	X	X
Socially Desirable Responding*	.76	16-item Balanced Inventory of Desirable Responding Short Form (BIDR-16) includes two subscales for measuring socially desirable responding: Self-Deceptive Enhancement and Impression Management *Source*: Hart et al. (2015) *Response options*: 7-point Likert scale from 1 = *not true* to 7 = *very true* *Example item*: "I am a completely rational person."	X	X	X	X
Demographics						
Age		"What is your age?"	X	X	X	X
Gender		"What is your gender?" (*Male / Female / Trans or non-binary*)	X	X	X	X
Semesters completed		"How many semesters of college have you completed?"	X	X	X	X
Race/ethnicity		"Which of the following best represents your racial and/or ethnic background?"	X	X	X	X

Instrument	Cronbach's α	Description	Uni A (W1)	Uni B (W1)	Prolific (W2)	Uni A (W3)
Employment		"Are you currently employed?" (*Yes / No*)	×	×	×	×
Generational status		"Are you first in your family to attend college?" (Yes / *No*)	×	×	×	×
Transfer status		"Are you a transfer student?" (*Yes / No*)	×	×	×	×
GPA		"What is the cumulative GPA from your coursework at [institution]?"	×	×	×	×
University name and location		"What is the name of the college/university you currently attend?"; "In which state is your college/university located?"			×	
Parental education level		"What is the highest level of education achieved by your [mother / father]?"	×	×	×	×
Social Class		"How would you describe your family's social class?" (*Poor / Working class / Lower-middle class / Middle class / Upper-middle class / Upper class*)	×	×	×	×
Income		"What is your current individual income?"; "What is your current annual income for your entire household?"	×		×	×
Subjective Socioeconomic Status		MacArthur Scale of Subjective Social Status (Adler et al., 2000); participants were presented with a picture of a ladder with 10 rungs and asked to select where they stand relative to others (in the U.S).	×		×	×

Note. W1 = Wave 1, Spring 2020 (Mar 3 – May 31); W2 = Wave 1, Summer 2020 (May 20–May 21); W3 = Wave 3, Summer/Fall 2020 (Jun 19–Dec 2); SES = Socioeconomic Status. Instruments/items collected on behalf of the collaborating research team are denoted by '*'. Instruments that were analyzed using item response models are denoted by '†'; all other items/instruments were transformed into mean composites.

Results

Table 3 presents the differences between the three subsamples on key demographic and psychological variables. The online sample was older but comparable in social class and GPA to University A. University B students had lower GPAs and higher social class than the other two subsamples. University A students reflected a slightly more positive attitude towards help-seeking but also a stronger perception of competitiveness among students. Students from Universities also reported higher commitment to the university and lower intentions to drop out. University B students reported slightly lower growth mindset and perceived faculty growth mindset and higher dropout intentions. The three samples are comparable in terms of self-efficacy, impostor phenomenon, sense of belonging, and competence.

Primary Results

As expected, perceiving instructors at one's institution to hold a malleable view of intelligence was correlated, on average, with more positive attitudes toward academic help-seeking, controlling for students' own mindset and the experience of academic difficulty in the previous month, $\chi^2(1, 1485) = 52.3$, $p < .001$, $r_{partial} = .18$. (See Table 4 for standardized regression estimates and associated standard errors.) Not facing academic difficulty was associated with a more positive attitude toward help-seeking.

Evaluating the same model with impostor phenomenon as the dependent variable, we find that although faculty growth mindset negatively predicted students' impostor feelings ($\beta = -0.09$, $SE = 0.03$), the effect was smaller than for other covariates, all of which significantly predicted impostor feelings except GPA. Students higher on growth mindset reported lower impostor phenomenon, as did students not facing academic difficulty. Conversely, students who did face academic difficulty in the previous month were more likely to report feeling like an imposter. Model comparison with a null model that did not include faculty growth mindset indicated that faculty growth mindset did add predictive power above and beyond the other covariates, $\chi^2(1, 1485) = 11.9$, $p < .001$, $r_{partial} = -.09$. Identical analyses with belonging uncertainty and competitiveness as outcomes show that faculty growth mindset was associated with lower belonging uncertainty ($\chi^2(1, 1485) = 57$, $p < .001$, $r_{partial} = -.19$), as well as lower perceptions of competitiveness ($\chi^2(1, 1485) = 172$, $p < .001$, adjusted $r_{partial} = -.32$).

In multiple regression models that included students' growth mindset, GPA, and academic difficulty as covariates, perceptions of competitiveness did not predict attitude toward help-seeking, $\beta = -0.03$, $SE = .03$, $\chi^2(1, 1531) = 1.59$, $p = .21$, $r_{partial} = -.02$, but higher feelings of impostor phenomenon did, $\beta = -0.29$, $SE = 0.02$, $\chi^2(1, 1533) = 138$, $p < .001$, $r_{partial} = -.29$. In this sample, students' own growth mindset was indeed correlated moderately with perceived faculty growth mindset, $r(1660) = .30$, $p < .001$, CI [0.25, 0.34], and students higher in growth mindset were less likely to view help-seeking as threatening, $r(1655) = -0.15$, $p < .001$ CI [-0.19, -0.10].

Table 3

Baseline Differences Between Subsamples in Study 1

Variable	University A *M* (*SD*)	University B *M* (*SD*)	Prolific *M* (*SD*)	*F* (*df*1, *df*2)	*p*
Demographic Variables					
Age (years)	20.75 (3.0)	20.97 (3.4)	23.08 (6.2)	48.5 (2, 1640)	<.001
Social Class (1-6 scale)	3.96 (1.21)	4.25 (0.92)	3.71 (1.13)	23.5 (2, 1635)	<.001
Semesters Completed	4.68 (2.23)	4.51 (2.72)	4.93 (2.51)	3.27 (2, 1627)	.04
GPA (1-4 scale)	3.47 (0.40)	3.13 (0.59)	3.48 (0.59)	74.4 (2, 1564)	<.001
Psychological Variables					
Help-Seeking (MSLQ)	4.08 (1.13)	3.88 (1.23)	3.95 (1.31)	3.64 (2, 1656)	.03
Help-Seeking as Threatening	2.36 (1.01)	2.24 (0.99)	2.25 (1.06)	2.57 (2, 1654)	.08
Perceptions of Help-Seeking Behavior	5.17 (0.75)	4.87 (0.75)	5.08 (0.79)	20 (2, 1652)	<.001
Self-Perception of Help-Seeking Behavior	3.39 (0.89)	3.26 (0.84)	3.27 (0.99)	3.79 (2, 1654)	.02
Help-seeking Behavior	1.98 (0.60)	1.98 (0.61)	–	0.14 (1, 624)	0.71
Self-Help	5.05 (1.04)	–	5.21 (0.96)	6.65 (1, 1031)	.01
Perceived Faculty Growth Mindset	3.94 (1.07)	3.54 (1.19)	4.14 (0.98)	29.8 (2, 1592)	<.001
Growth Mindset	4.16 (1.12)	4.00 (1.08)	4.21 (1.17)	4.03 (2, 1654)	.02
Self-Efficacy	7.59 (1.70)	–	7.52 (1.83)	0.44 (1,1028)	.51
Imposter Feelings	3.32 (0.67)	3.29 (0.60)	3.30 (0.71)	0.35 (2, 1656)	.70
Belonging	4.96 (1.41)	4.84 (1.32)	4.93 (1.56)	0.85 (2, 1598)	.43
Belonging Uncertainty	2.42 (1.01)	2.54 (1.04)	2.48 (1.10)	1.80 (2, 1654)	.17
Autonomy (SDT)	4.60 (0.88)	4.58 (0.88)	4.60 (0.99)	0.06 (2, 1654)	.94
Competence (SDT)	4.51 (1.02)	4.38 (0.92)	4.54 (1.11)	2.74 (2, 1653)	.06
Relatedness (SDT)	5.25 (1.00)	5.20 (0.99)	4.97 (1.03)	11.9 (2, 1654)	<.001
Perceptions of Competitiveness	4.87 (1.30)	4.27 (1.30)	3.54 (1.43)	151 (2, 1651)	<.001
University Commitment	4.90 (0.90)	4.64 (1.02)	4.43 (1.18)	33.6 (2, 1652)	<.001
Dropout Intentions	1.70 (0.94)	1.92 (1.06)	1.84 (1.03)	6.62 (2, 1589)	<.05

Note. GPA = Grade Point Average; MSLQ = Motivated Strategies Learning Questionnaire; SDT = Self-Determination Theory

Table 4

Standardized Regression Estimates for Primary Results in Study 1

	Help-Seeking Composite		Impostor Phenomenon		Belonging Uncertainty		Competitiveness	
	β	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Faculty Growth Mindset	0.19***	0.03	-0.09***	0.03	-0.19***	0.03	-0.34***	0.03
Growth Mindset	0.14***	0.03	-0.11***	0.03	-0.04	0.03	0.01	0.03
GPA	0.13***	0.03	0.04	0.03	-0.15***	0.03	0.01	0.03
Academic Difficulty (No)	0.08	0.05	-0.28***	0.05	-0.29***	0.04	-0.19***	0.04
Academic Difficulty (Yes)	-0.05	0.03	-0.13***	0.03	0.12***	0.03	0.09**	0.03
Adjusted R^2	.10		.06		.13		.13	

Note. GPA = Grade Point Average. Intercept terms have been removed. ***p* < .01. ****p* < .001.

Secondary Results

As a descriptive measure, we first present a correlation matrix (Table 5) that depicts zero-order correlations among all the key variables. As expected, especially given the large sample size, most variables are significantly correlated with one another. Next, in Table 6 and Figure 3, we provide standardized regression estimates from general linear (OLS) models predicting the five help-seeking instruments used in the composite measure from all the key variables. Controlling for all other covariates, Faculty Growth Mindset positively predicted Perceptions of Help-Seeking, and negatively predicted perceiving Help-Seeking as Threatening. Help-Seeking Behavior was positively associated with Social Fit and Positive Affect and negatively with Autonomy and *not* facing Academic Difficulty in the previous month (that is, students who reported a higher sense of autonomy and did not experience academic difficulty in the previous month were less likely to report seeking help).

Results of regression models predicting these outcomes, along with the Help-Seeking Composite, from demographic variables are available in Appendix B (Table B1). GPA emerged as the strongest predictor and is positively related to perceptions of help-seeking but negatively related to actual help-seeking behavior. This seems in line with the expectation that high-achieving students would be less inclined to seek help (since they don’t need it), but they likely view help-seeking as an adaptive self-regulatory academic strategy. Older students were more likely to perceive help-seeking as less threatening but were less likely to seek help. Other demographic characteristics like gender, underrepresented minority status, first-generation college student status, transfer status, and social class do not show consistent patterns, except that being female and underrepresented minority predicts more positive perceptions of help-seeking in general. These null and slightly positive results counter our expectations and the existing literature; although they should be interpreted in light of the sample makeup, if true, they paint a more optimistic picture—if students’ attitude toward help-seeking is less dependent on immutable demographic characteristics, that leaves more room for global interventions that can help all students.

Table 5

Intercorrelations for Study 1 Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Help-Seeking																			
2. Faculty Growth Mindset	.23																		
3. Growth Mindset	.18	.30																	
4. Self-Efficacy	.43	.24	.14																
5. Impostor Feelings	-.30	-.14	-.12	-.35															
6. Belonging	.34	.26	.15	.37	-.32														
7. Belonging Uncertainty	-.35	-.23	-.08	-.43	.46	-.69													
8. Autonomy	.32	.24	.23	.41	-.44	-.46													
9. Competence	.41	.22	.20	.51	-.54	.45	-.51	.65											
10. Relatedness	.37	.24	.20	.33	-.26	.44	-.42	.62	.55										
11. Competitive	-.05	-.34	-.09	-.10	.12	-.13	.11	-.13	-.10	.01									
12. University Commitment	.25	.26	.11	.28	-.18	.55	-.49	.32	.31	.40	.01								
13. Dropout Intentions	-.22	-.22	-.05	-.34	.24	-.40	.51	-.27	-.32	-.26	.06	-.60							
14. Academic Difficulty	-.10	-.09	.03	-.28	.18	-.14	.25	-.16	-.22	-.07	.15	-.09	.22						
15. Self-Concept Fit	.33	.30	.16	.35	-.20	.64	-.46	.42	.40	.42	-.14	.54	-.32	-.08					
16. Goal Fit	.37	.37	.16	.46	-.23	.59	-.47	.45	.45	.39	-.23	.47	-.37	-.16	.83				
17. Social Fit	.34	.28	.15	.33	-.28	.56	-.46	.49	.41	.48	-.17	.42	-.27	-.10	.79	.74			
18. State Authenticity	.36	.23	.16	.42	-.31	.59	-.47	.47	.46	.42	-.15	.44	-.32	-.12	.61	.57	.59		
19. RSOS	.31	.17	.14	.33	-.25	.48	-.41	.43	.40	.40	-.06	.41	-.29	-.12	.52	.46	.54	.60	

Note. RSOS = Real Self Observed Self. Help-Seeking is the composite of 13 items. Non-significant correlations are presented in grey.

Table 6

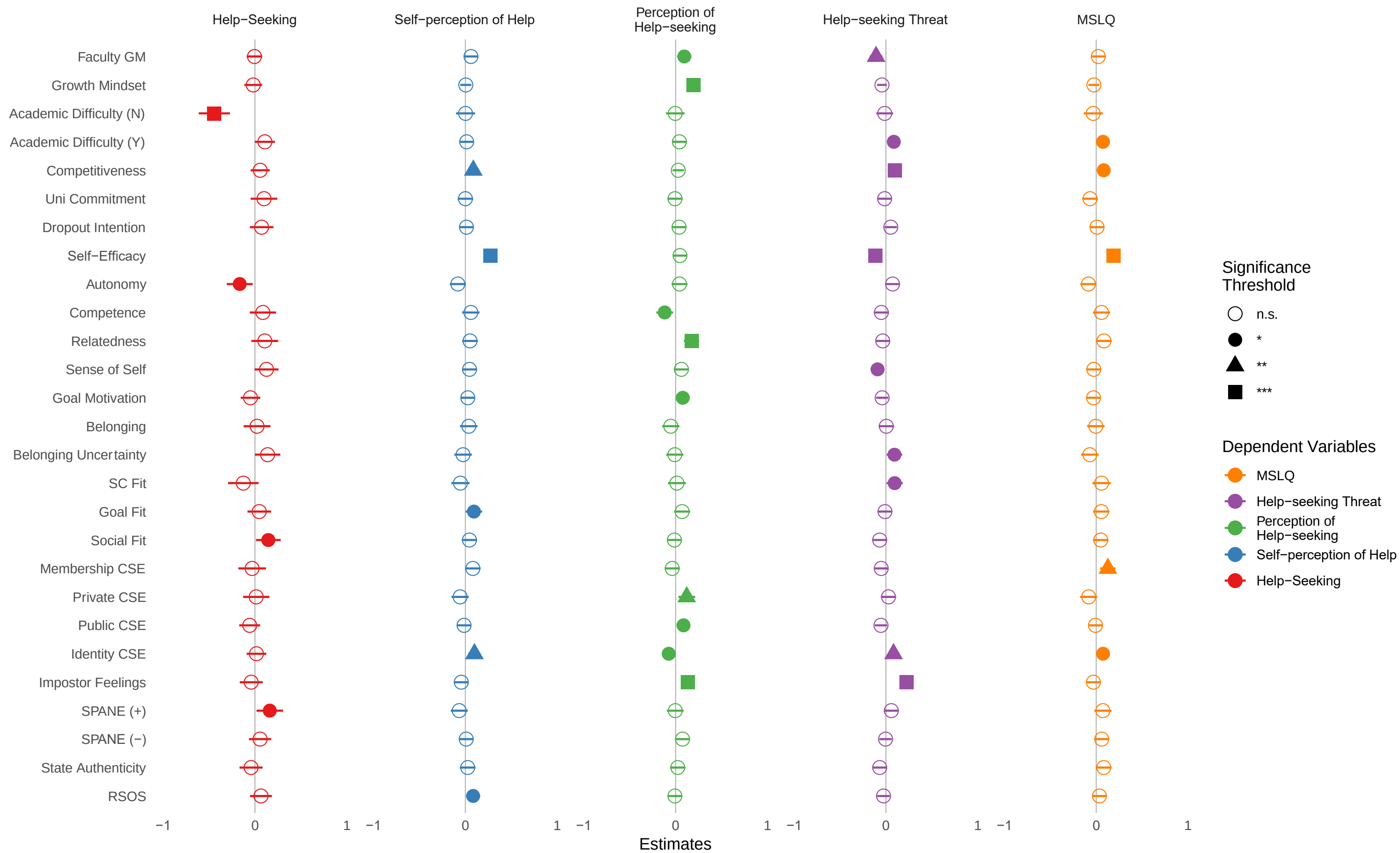
Regression Estimates from Models Predicting Help-Seeking Variables

Variable	Help-Seeking Behavior		Self-Perception of Help		Perception of HS		HS as Threatening		MSLQ	
	β	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Faculty Growth Mindset	-0.01	0.04	0.06	0.04	0.09*	0.04	-0.11**	0.04	0.02	0.04
Growth Mindset	-0.02	0.05	0.01	0.03	0.20***	0.04	-0.04	0.03	-0.03	0.03
Academic Difficulty (N)	-0.44***	0.09	0.00	0.05	0.00	0.04	-0.01	0.05	-0.03	0.05
Academic Difficulty (Y)	0.11	0.06	0.01	0.04	0.04	0.04	0.09*	0.03	0.07*	0.04
Competitiveness	0.06	0.05	0.09*	0.03	0.03	0.04	0.10***	0.03	0.08*	0.03
University Commitment	0.10	0.07	0.00	0.04	-0.01	0.04	-0.01	0.04	-0.07	0.04
Dropout Intention	0.07	0.07	0.01	0.04	0.04	0.04	0.05	0.03	0.01	0.04
Self-Efficacy			0.27***	0.04	0.05	0.04	-0.12***	0.03	0.19***	0.04
Autonomy	-0.17*	0.07	-0.08	0.04	0.04	0.04	0.07	0.04	-0.09	0.04
Competence	0.09	0.07	0.06	0.05	-0.12*	0.04	-0.05	0.04	0.06	0.05
Relatedness	0.11	0.07	0.05	0.04	0.17***	0.04	-0.04	0.04	0.08	0.04
Sense of Self	0.13	0.07	0.05	0.04	0.06	0.04	-0.09*	0.04	-0.03	0.04
Goal Motivation	-0.05	0.05	0.03	0.04	0.08*	0.04	-0.04	0.03	-0.03	0.04
Belonging	0.02	0.07	0.04	0.05	-0.05	0.04	0.00	0.04	0.00	0.05
Belonging Uncertainty	0.14	0.07	-0.02	0.05	-0.01	0.04	0.09*	0.04	-0.07	0.05
Self-Concept Fit	-0.13	0.08	-0.05	0.05	0.01	0.04	0.10*	0.04	0.06	0.05
Goal Fit	0.05	0.07	0.09*	0.04	0.07	0.04	-0.01	0.04	0.06	0.05
Social Fit	0.15*	0.07	0.04	0.04	-0.01	0.04	-0.07	0.04	0.05	0.04
Membership CSE	-0.03	0.08	0.08	0.04	-0.04	0.04	-0.05	0.04	0.13**	0.04
Private CSE	0.01	0.07	-0.06	0.05	0.12**	0.04	0.03	0.04	-0.08	0.05
Public CSE	-0.06	0.06	-0.01	0.04	0.09*	0.04	-0.05	0.03	-0.01	0.04
Identity CSE	0.02	0.05	0.10**	0.03	-0.07*	0.04	0.08	0.03	0.07*	0.03
Clance IP	-0.04	0.06	-0.04	0.04	0.13***	0.04	0.22***	0.03	-0.03	0.04
Positive Affect	0.16	0.07	-0.07	0.05	0.00	0.04	0.06	0.04	0.07	0.05
Negative Affect	0.06	0.06	0.01	0.04	0.08	0.04	0.00	0.03	0.06	0.04
State Authenticity	-0.04	0.06	0.03	0.04	0.02	0.04	-0.07	0.04	0.08	0.04
RSOS	0.07	0.06	0.09*	0.04	-0.01	0.04	-0.03	0.03	0.03	0.04
Adjusted R^2	.10		.23		.18		.37		.19	

Note. Intercept terms have been removed. CSE = Collective Self-Esteem; IP = Impostor Phenomenon; RSOS = Real Self Observed Self.
*p < .05. **p < .01. ***p < .001.

Figure 3

Regression Estimates from Models Predicting Help-Seeking



Note. MSLQ = Motivated Strategies Learning Questionnaire; CSE = Collective Self-Esteem; RSOS = Real Self Observed Self; SPANE = Survey of Positive and Negative Experience. Self-efficacy and Help-Seeking Behavior were not measured together, hence the missing estimate.

Specification Curve Analysis

The next exploratory step was to test the predictive power of the primary independent variable, Faculty Growth Mindset, in predicting outcomes related to students' psychological experience. Furthermore, although our primary independent and dependent variables were selected a priori, the correlational nature of our data and the relatively large number of variables measured could raise legitimate concerns regarding selective reporting. To preemptively address these concerns, we conducted a specification curve analysis (Simonsohn et al., 2020). This analysis takes a kitchen-sink approach such that all 'reasonable,' non-redundant, and theoretically valid models are fit to data simultaneously, and results are presented graphically for interpretation. Although this is an atypical use of specification curve analysis, which is generally used to analyze different specifications of the same model, we deem this a useful exploratory exercise. If Faculty Growth Mindset does not correlate with any psychological variables deemed important for academic motivation and does not predict any aspects of attitude toward help-seeking, this line of investigation may not be worth pursuing in future studies.

To assess the robustness of the effect of Faculty Growth Mindset in predicting outcomes of interest, we conducted an analysis of all possible model specifications that included Faculty Growth Mindset as the predictor and all instruments/items measuring Help-Seeking (including the 13-item composite) and all other key dependent variables as outcomes. Most outcomes were expected to be positively correlated with Faculty Growth Mindset, but some were expected to show a negative relationship (e.g., Impostor Phenomenon, Belonging Uncertainty). In our sample, Faculty Growth Mindset does not differ significantly by gender, race/ethnicity, age, first-generation status, or any other demographic variable. Consequently, we included as controls three covariates that were correlated with Faculty Growth Mindset in the main analysis: Grade Point Average (GPA), Growth Mindset (GM), and Academic Difficulty (AD).

These decisions yielded 248 model specifications (eight specifications each for 31 outcome variables; one with no controls, one with all controls, three with a single control, and the remaining three included pairs of controls). The top panel in Figure 5 presents the specification curve, which plots the standardized regression estimates for all 248 models, ordered from smallest to largest. Estimates ranged from -0.34 (the left-most estimate) to 0.37 (the right-most estimate). Negative estimates are depicted in red, statistically non-significant estimates in gray, and positive estimates in blue. For linking regression estimates to model specifications, the middle panel in Figure 5 displays the dependent variable and controls included in each specification. (All models included Faculty Growth Mindset as the predictor.) Sample sizes are shown in the bottom panel; sample sizes are smaller for a subset of models because some variables were not measured during all waves of data collection.

The model with the largest negative estimate included perceptions of Competitiveness as the dependent variable and Growth Mindset as a control. All models with Competitiveness as the dependent variable result in similar estimates, regardless of the controls included in the model. In other words, the specification curve analysis supports the idea that perceived Faculty Growth Mindset is negatively associated with perceptions of Competitiveness at the university, and this effect is robust to different choices of controls. Similarly, models that produce the largest estimates, all related to the Goal Fit (feeling that the university environment supports one's personal goals), are also robust across choices of covariates.

Higher dispersion of estimates of an outcome indicates that the effect of Faculty Growth Mindset on that outcome is impacted by the choice of controls. In the analysis, several outcomes (SPANE, Imposter Phenomenon, Importance to Identify (CSE), MSLQ, Self-Help, and HS behavior) do not appear to be affected meaningfully by the choice of controls. By contrast, as in the case of Autonomy and Competence, the estimates are smaller if the model includes all three controls and larger without any controls. The specification curve analysis thus provides an unbiased exploratory overview of the variation of estimates depending on model specification. Note that the dispersion of the estimates across specifications is relative to the proximity of estimates similar in magnitude, which differs across specifications.

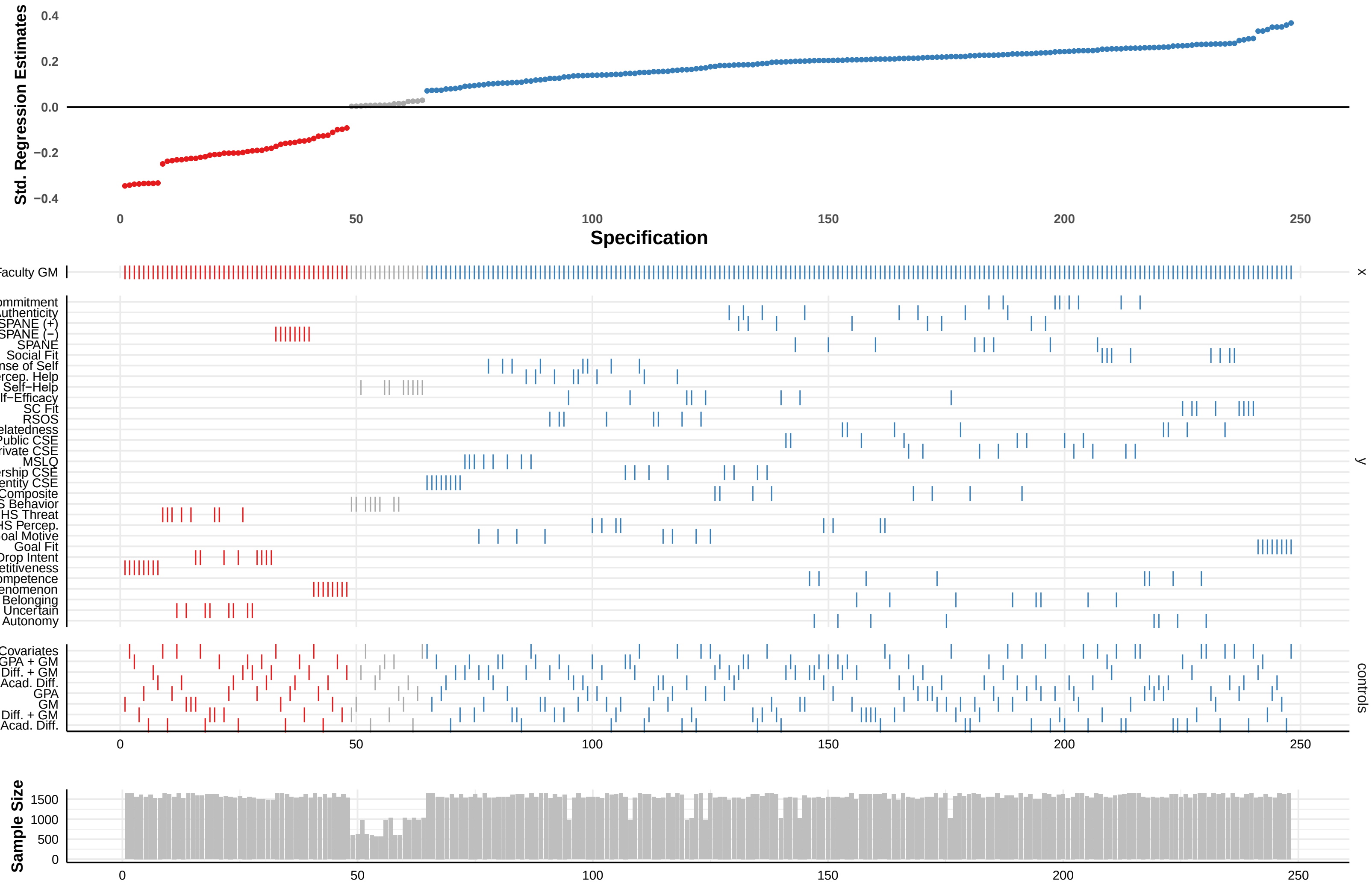
Figure 5 plots, in ascending order, the R^2 values and the associated regression estimates for all models. R^2 values for models that are non-significant were non-trivial because the control variables explain a portion of the variance in the outcomes. Regression estimates that are non-significant are shown in gray (with 95% confidence intervals), those that are significant and in the expected direction are colored blue, and since no significant effects were in the unexpected direction, there are none in red. The median regression estimate (standardized) was 0.17 (mean = 0.11). All except 16 models (8%) were significant at the $\alpha = .05$ threshold. Only two outcome variables were not significantly correlated with Faculty Growth Mindset, the single Self-Help item that was excluded from the composite and self-reported Help-Seeking Behavior.

When we subset the data based on attention checks (those who passed, those who failed, and everyone: $248 \times 3 = 744$ specifications), results remained the same (median $\beta = 0.17$, min $\beta = -0.37$, max $\beta = 0.41$). In addition to the 16 null results from the whole sample, there were 38 non-significant estimates from the subset of participants who failed the three attention checks and 24 from the subset that passed (10% of the total). When we included all other demographic variables as controls, the number of specifications increased to 126,976, and the results did not change (median $\beta = 0.18$, min $\beta = -0.38$, max $\beta = 0.37$). 7% of the models (8,269) were non-significant and consisted of the following outcome variables: Self-Help, Help-Seeking Behavior, and Importance to Identity (Collective Self-Esteem).

Thus, it appears that Faculty Growth Mindset explains comparable variance in many of the variables that tap into students' psychological experience in college, and these effects seem robust to different specifications. Faculty Growth Mindset also significantly predicted perceptions of and feelings toward academic Help-Seeking. Crucially, however, we consistently fail to see an effect when actual help-seeking behavior is concerned. Although models predicting help-seeking behavior are tested with lower power (since Help-Seeking Behavior was only measured during the first wave of data collection), this is an important finding that we discuss further in the Discussion.

Figure 4

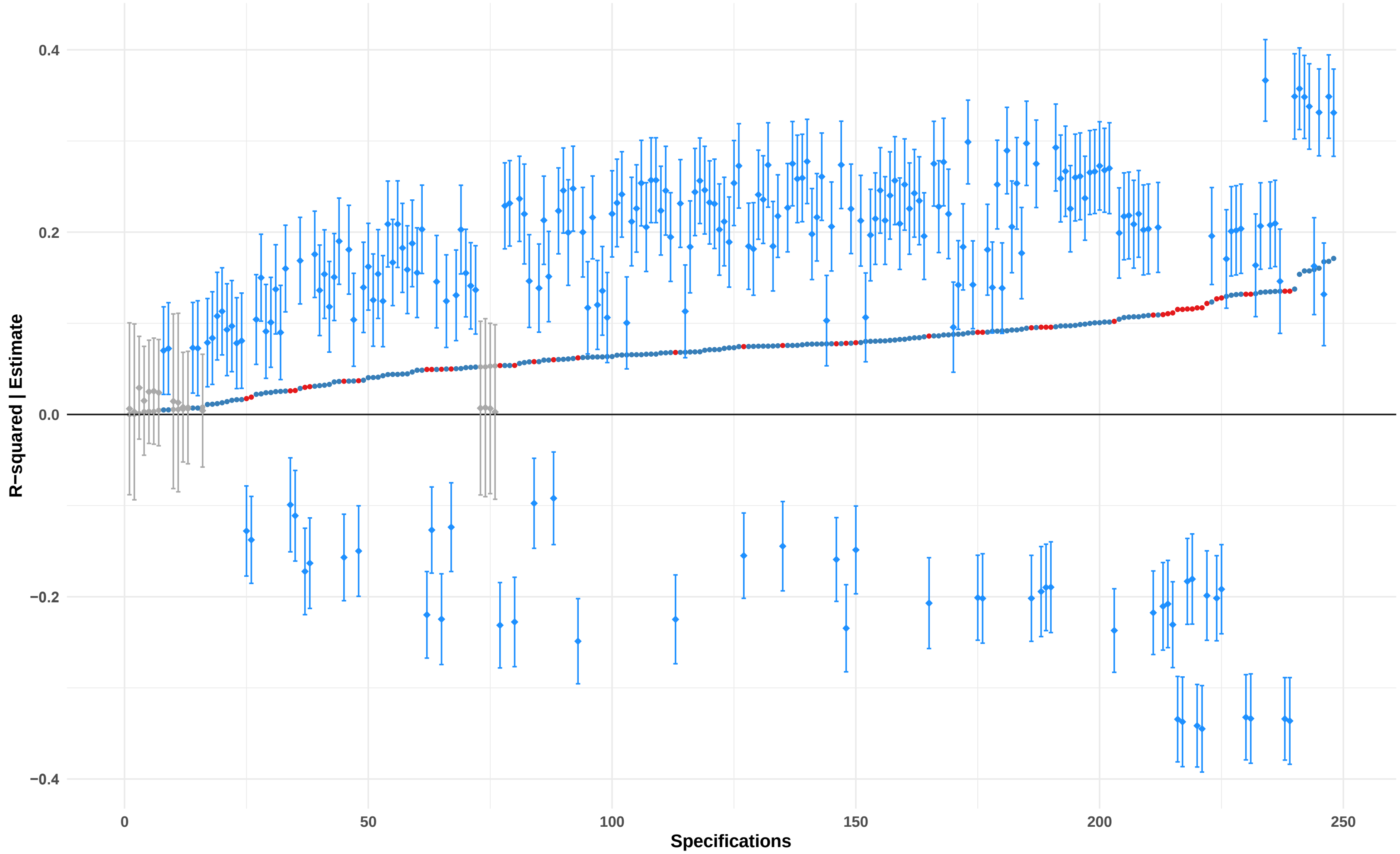
Specification Curve, Model Choices, and Sample Sizes



Note. GM = Growth Mindset; MSLQ = Motivated Strategies Learning Questionnaire; SC = Self-Concept; CSE = Collective Self-Esteem. Specification curve (top) plotted with sample sizes (bottom). Middle panel represent model choices, with independent variable at the top, dependent variables in the middle, and controls at the bottom. Estimates are in ordered by sign and magnitude.

Figure 5

Regression Estimates and R^2 Values for Specification Curve Analysis



Note. R^2 values (center; ordered by magnitude) and standardized regression estimates (with error bars) plotted against specification numbers. Dots for R^2 values are blue for positive estimates, and red for negative estimates (indicated by the estimates and confidence intervals below the zero line). Regression estimates in gray represent non-significant effects; blue represents effects in the expected direction, and red, in the unexpected direction.

Regularized Regression

The primary goal of the current study was to assess whether attitude toward help-seeking, the main dependent variable, correlates with students' perceptions of their instructors' growth mindset, the primary independent variable. In the previous section, our aim was to ascertain the predictive power of the independent variable for predicting all outcome variables measured in this study. As a last step in the auxiliary analysis, we mirror this approach with the dependent variable and seek to determine the strongest predictors of attitude toward help-seeking. The OLS regressions presented in Table 6 (help-seeking instruments predicted by all outcome variables) shed some light on this question, but the large number of estimated coefficients preclude a clear interpretation. Therefore, to assist in interpretation and help increase prediction accuracy, we estimate the linear model with shrinkage and try to limit the coefficients to those that explain the most variance. To that end, we use regularized regressions that apply L_1 and L_2-norm penalties and use cross-validated R^2 values for model comparison.

One of the main disadvantages of complex models with multiple predictors is the risk of overfitting. The resulting models likely explain the data well (low bias) but are poor at prediction (high variance). Although our study is well-powered, our predictors are highly correlated, which can also result in unreliable estimates. One solution is the use of the 'least absolute shrinkage and selection operator' (lasso) regression, which reduces variance by adding bias (Tibshirani, 1996). This is done through regularization, where we set a prior distribution and penalize coefficients that are too small (or too large) and shrink some to zero. The shrinkage is based on a parameter, lambda (λ), which is evaluated at different levels to find a value that optimally minimizes the error. In lasso regularization, the prior is a double exponential, or Laplacian, distribution (Tibshirani, 1996; Tibshirani et al., 2012). Ridge regression is another version of the same method that assumes a Gaussian prior. Ridge regression results in smoother shrinkage that depends on the strength of the correlation among predictions; the model penalizes (shrinks) the coefficients of predictors that are correlated but retains all predictors in the final model. On the other hand, a lasso penalty retains one from pairs of predictors that are highly correlated (Friedman et al., 2010).

The model solves the following equation:

$$min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N} l(y_i, \beta_0 + \boldsymbol{\beta^T x_i}) + \lambda\left[(1-\alpha)\|\beta\|_{l_2}^2/2 + \alpha\|\beta\|_{l_1}\right],$$

where $\boldsymbol{x_i}$ and y_i are the predictor and response variables for i = 1, 2… N, and β_0 *and* $\boldsymbol{\beta^T}$ are the model coefficients; $l(y_i, \eta_i)$ is the negative log-likelihood for each observation i, which in the case of a standard OLS regression is $\frac{1}{2}(y_i - \eta_i)^2$. λ is the tuning parameter that determines the amount of shrinkage and α determines the nature of the penalty; when α = 1, the first expression inside the square brackets is 0 and results in a lasso penalty, which uses the L_1-norm (sum of absolute values of βs). When α = 0, the second expression is 0, and we get the ridge penalty, which uses the L_2-norm (sum of squared βs).

For this analysis, we first derive estimates from a cross-validated OLS model for comparison; the model includes the 13-item Help-Seeking composite as the outcome variable and all key variables (n = 26) as predictors. Then, the regularized regression models are evaluated using penalized maximum likelihood and the regularization or tuning parameter, lambda (λ), which determines the strength of the penalty (using the *glmnet* package in R;

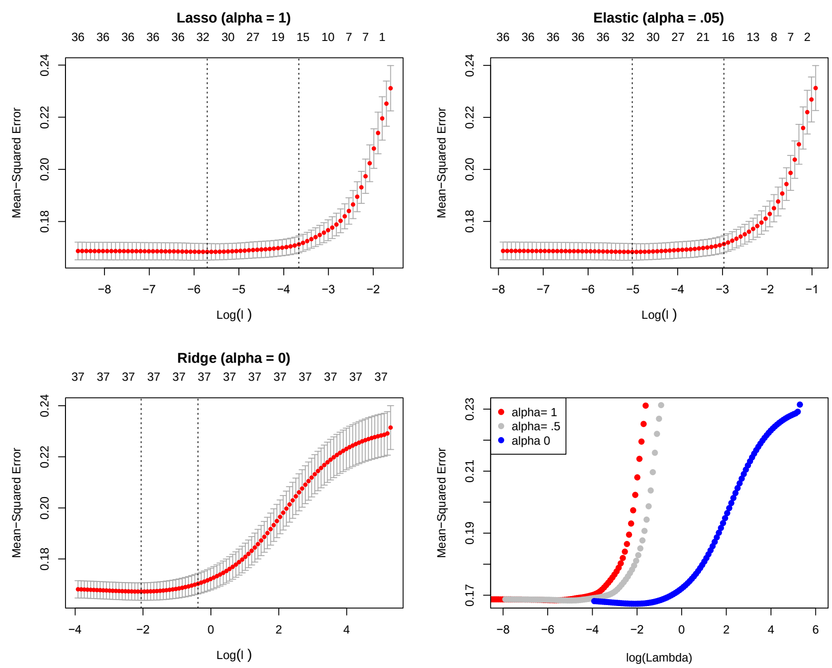
Friedman et al., 2009, 2021). These models are also cross-validated (coefficients and errors are derived from a different subset of the data).

Figure 6 shows the results of a 10-fold cross-validation (Efron & Tibshirani, 1993) used to identify the λ that yields the lowest prediction error for both lasso and ridge regressions, as well as an elastic net, which uses a penalty that combines L_1 and L_2 (Zou & Hastie, 2005). The left dotted lines in each panel reflect the λ that results in the lowest mean square errors. After finding the appropriate λs, we fit lasso and ridge regressions using single-fold cross-validation. Cross-validated coefficients of determination (R^2_{cv}) for model comparison were derived using 10-fold cross-validation (See Figure 7). Table 7 compares the coefficients from the three different models. Standard errors were calculated using the *selectiveInference* package in R (Tibshirani et al., 2019). However, we recommend caution while interpreting these standard errors; although methods have been developed in recent years, standard errors are less reliable for biased estimates that result from penalized estimation methods (Tibshirani, 1996).

Both ridge and lasso resulted in similar coefficients of determination, but ridge regression has been shown to perform better for many small effects (Tibshirani, 1996), which is true in our case. However, since our goal was primarily to determine the strongest predictors of help-seeking behavior, a more parsimonious model that is easier to interpret (i.e., the lasso regression) seems more appropriate.

Figure 6

Cross-Validation Results for Shrinkage Parameter Selection



Note. The mean square error as a function of the log of the 10-fold cross-validated shrinkage parameters (λ) for lasso (top left) and ridge (bottom left), and elastic net (top right). Bottom right panel shows the relative performance of all three.

Table 7

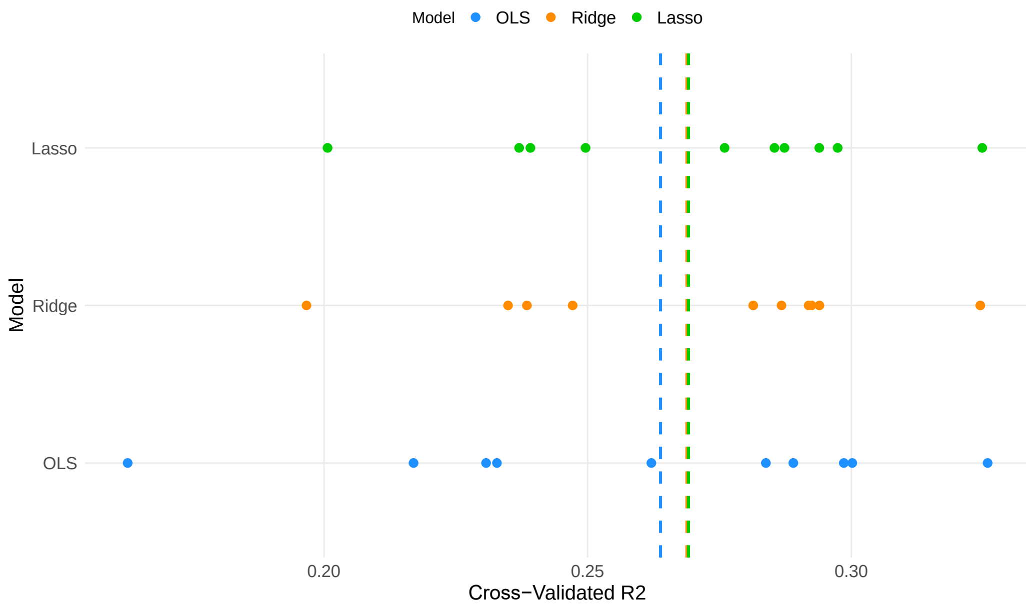
Standardized Estimates for OLS, Ridge, and Lasso Regressions

Predictor	OLS Estimates		Ridge Estimates		Lasso Estimates	
	β	*SE*	*β*	*SE*	*β*	*SE*
Faculty Growth Mindset	0.06	0.02	0.06	0.02	0.05	0.02
Growth Mindset	0.05	0.01	0.03	0.02	0.03	0.02
Self-Efficacy	0.12	0.02	0.10	0.02	0.10	0.02
Impostor Phenomenon	-0.06	0.02	-0.05	0.03	-0.05	0.02
Belonging	-0.01	0.02	0.00	0.02	.	
Belonging Uncertainty	-0.04	0.02	-0.03	0.02	-0.02	0.02
Autonomy	-0.06	0.02	-0.05	0.03	-0.04	0.02
Competence	0.07	0.02	0.08	0.02	0.09	0.02
Relatedness	0.07	0.02	0.06	0.02	0.06	0.02
Competitiveness	0.03	0.02	0.02	0.03	0.01	0.02
University Commitment	-0.03	0.02	-0.02	0.02	-0.01	0.03
Dropout Intention	0.00	0.02	0.01	0.04	0.01	0.05
Academic Difficulty	0.00	0.01	0.00	0.04	.	
Sense of Self	0.01	0.02	0.02	0.01	0.01	0.05
Goal Motivation	0.01	0.02	0.01	0.04	0.01	0.05
Self-Concept Fit	-0.01	0.02	-0.02	0.01	.	
Goal Fit	0.08	0.02	0.08	0.02	0.07	0.02
Social Fit	0.01	0.02	0.02	0.03	0.01	0.04
Membership CSE	0.04	0.02	0.04	0.02	0.03	0.03
Private CSE	-0.02	0.02	-0.01	0.01	.	
Public CSE	0.01	0.02	0.01	0.01	.	
Identity CSE	0.02	0.02	0.02	0.02	0.01	0.04
Positive Affect	0.01	0.02	0.02	0.04	0.01	0.06
Negative Affect	0.03	0.02	0.03	0.01	0.02	0.02
State Authenticity	0.05	0.02	0.05	0.01	0.05	0.02
RSOS	0.03	0.02	0.02	0.03	0.02	0.03
R^2_{cv}	0.26		0.27		0.27	

Note. OLS = Ordinary Least Squares; CSE = Collective Self-Esteem; RSOS = Real Self Observed Self.

Figure 7

Cross-Validated R^2 values for OLS, Ridge, and Lasso Regressions



Note. Dotted lines represent the average of 10 cross-validated R^2 values.

We see that estimates from all three models are relatively similar. The lasso penalty admits 21 non-zero coefficients, although many of them are close to the smallest effect size we were powered to detect (r = .01). High-quality, randomized trials in education rarely yield effect sizes larger than .20, and studies with sample sizes 2000 or higher are estimated to have an average effect size closer to .10 (Cheung & Slavin, 2016). Using .10 as a heuristic criterion, we see that only a few predictors result in estimates close to that number: Self-Efficacy, Competence, and Self-Concept Fit. Faculty Growth Mindset has an estimate of .05, which is small. However, Faculty Growth Mindset predicts outcomes above and beyond a number of important covariates. Given that Faculty Growth Mindset might be more amenable to instructors' control, we see this result as relatively promising.

Discussion

The aim of this investigation was to unearth an association between academic help-seeking behavior and students' perceptions of their instructors' implicit beliefs about intelligence if such a relationship existed. We carried out a high-powered observational (non-experimental) study with college students and measured several psychological factors associated with academic success and well-being. We found that college students' perceptions of their instructor's implicit beliefs about intelligence robustly correlate with multiple indicators of student experience, including students' attitude toward help-seeking, impostor feelings, and perceptions of competitiveness. Some of the strongest negative relationships we observe are between faculty growth mindset and perceptions of the competitiveness of the college environment, impostor feelings, belonging uncertainty, and perceiving help-seeking as threatening. That is, perceiving faculty at their university to hold a malleable view of intelligence is associated with lower perceptions of competitiveness, lower impostor feelings, lower belonging uncertainty, and a lower likelihood of viewing help-seeking as threatening. Conversely, we see positive associations with State Authenticity as Fit (SAFE; self-concept, goal, and social fit) and Self-Determination measures (autonomy, relatedness, and competence). Importantly, the association between faculty growth mindset and attitude toward help-seeking, albeit modest in size, holds even when we account for all other factors measured in this study.

Results from the regularized regressions indicate that help-seeking behavior is predicted most strongly by personal agency variables like self-efficacy and competence, as well as a lack of academic difficulty in the preceding month. It appears that students who are more high-achieving and confident in their abilities feel more positively about seeking help but also seek help less, which is to be expected. Although we find relationships among perceived faculty growth mindset and cognitive and psychological help-seeking variables, we do not find consistent effects when it comes to actual help-seeking behavior, which seems influenced primarily by students' experience of academic difficulty. We should note, however, that the data on actual help-seeking behavior were collected during 2020 at the peak of a global pandemic, which may have affected both the propensity and the affordances to engage in behaviors that constitute our operationalization (asking questions, visiting office hours, or seek help from peers). This is a major limitation of the study, and we present other limitations in the next section.

Limitations and Future Directions

The observational nature of our data does not allow us to make causal claims about the relationships we observe between faculty growth mindset and students' psychological experiences on college campuses. A spillover effect may be quite likely, such that students who are thriving in their college environment are more likely to perceive and rate faculty more positively. The large number of correlated variables collected in this study additionally complicates the interpretation of the findings. Finally, the study took place in 2020 during a global crisis that massively affected educational norms and behaviors. This is a major limitation of the study and should be taken into consideration when interpreting the results, especially in relation to academic help-seeking behavior.

Nevertheless, our results indicate a relationship between students' perceptions of their instructors' theories about intelligence and their attitude toward help-seeking, as well as factors that could affect help-seeking behavior, like perceptions of competitiveness of the college environment and feelings of impostorism. Thus, we aim to explore faculty growth mindset, or perceived implicit theories of intelligence, in future studies, focusing first on improved measurement and then on targeting relevant aspects of students' psychological experience and tracking their help-seeking behavior in specific courses.

Conclusion

The aim of this study was to test whether students' perceptions of their college instructors' theories of intelligence (faculty growth mindset) correlated with their attitude toward academic help-seeking. Although we do not observe a connection between perceived faculty growth mindset and actual help-seeking behavior (which may result from limitations due to data collection timing), we do find evidence to support the notion that perceiving instructors at one's university to hold a malleable view of intelligence is correlated with a more positive view of academic help-seeking, and with lower perceptions of competitiveness, impostor feelings, and belonging uncertainty. After observing a link between faculty growth mindset and academic help-seeking, the next logical step in the study of perceived theories of intelligence was the creation of a reliable and valid instrument that could be used to explore this construct further. In the next chapter, we describe the development and initial validation of an instrument for measuring perceived implicit theories of intelligence.

CHAPTER III

PERCIEVED IMPLICIT THEORIES OF INTELLIGENCE

Structured Abstract

Background: Measurement is the first and arguably the most crucial step in investigating causal phenomena. Although implicit theories of intelligence have been studied extensively, the instruments used to study them have not been studied well.

Purpose: To validate and investigate the psychometric properties of an instrument designed to measure students' perceptions of their instructors' implicit theories of intelligence (Perceived Implicit Theories of Intelligence; P-TOI)

Participants: 156 high school (n = 65) and college (n = 91) students recruited through word of mouth, an online data collection platform (Prolific; https://www.prolific.co), and a large-scale research network (Character Lab Research Network).

Research Design: Cross-sectional survey

Data Collection and Analysis: Data were collected between December 2020–January 2021 via Qualtrics (https://www.qualtrics.com) and analyzed using R (R Core Team, 2023), ACER Conquest (Adams et al., 2012), and BEAR Assessment System Software (BASS; Wilson & Sloane, 2000). Hypotheses and analysis plan were not preregistered.

Findings: The P-TOI instrument has adequate psychometric properties and shows sufficient internal consistency. Results support arguments for convergent and divergent validity, although there is some indication of differential item functioning for high school and non-U.S students.

Conclusion: The P-TOI instrument displays sufficient validity and reliability to warrant further study and use with slight modifications.

Perceived Implicit Theories of Intelligence

As alluded to in the introduction, educators' self-reported theories about intelligence have been shown to influence student outcomes (Canning et al., 2019), espcially in STEM (Science, Technology, Engineering, and Mathematics) education. Research has indicated that perceiving STEM college instructors to hold a fixed view of ability might result in greater psychological vulnerability, lower engagement, and lower performance (Muenks et al., 2020). The prevalence of the notion of the 'scientific genius' and the idea that brilliance is important for scientific success (Bian et al., 2018; Chestnut et al., 2018; Leslie et al., 2015; Meyer et al., 2015) places beliefs about the nature of intelligence at a crucial leverage point for influencing students' assessment of their aptitude for STEM disciplines.

If instructors' implicit theories bleed into the classroom through their interactions with students, they may affect students' perceptions of their potential to succeed. Crucially, what educators think and what their students believe they think may not always align. It is possible for instructors to hold the implicit belief that students are capable of significant intellectual growth but for it to not reflect in their behavior. Researchers have identified a recent proliferation of a *false growth mindset,* which results from believing in the malleability of intelligence, but behaving in a way that belies those beliefs (e.g., praising students' intelligence, telling students they 'just need to work hard'; Barger et al., 2022). Thus, it is vital to measure *perceived* theories of intelligence.

Perceived Theories of Intelligence (P-TOI) refer to students' perceptions of their instructors' beliefs about the malleability of intelligence (Kroeper et al., 2022; Muenks et al., 2020; Rattan et al., 2018). Rattan and colleagues (2018) define meta-lay theories as "individuals' beliefs about the lay theories that others hold regarding specific, relevant attributes." Relating these metatheories to intellectual potential, they state that "there are two ends of this continuum, ranging from a belief that not everyone has high intellectual potential (the nonuniversal belief), to the idea that almost everyone has high intellectual potential (the universal belief)" (p. 6). Muenks and colleagues (2020) similarly define it as "students' perceptions of their professors' mindset beliefs along the classic Dweckian dimension of the fixedness or malleability of intelligence" (p. 2).

Extant research has so far correlated instructors' theories of intelligence, as well as students' perceptions of these theories, with several psychological and academic outcomes (engagement, psychological experience, and performance; Fuesting et al., 2019; Muenks et al., 2020, 2021; Yeager et al., 2022). Canning et al. (2019) found that instructors' self-reported theories about intelligence were positively associated with student academic outcomes (official grades), and this effect was more substantial for students with racial/ethnic backgrounds under-represented in STEM. Controlling for confounding factors like previous achievement and faculty and course characteristics, students in courses taught by faculty members with a fixed mindset performed worse. Muenks and colleagues (2020) extended this line of research and assessed how college undergraduates' perceptions of their STEM instructors' theories about intelligence influenced their psychological experience in the course. Results show that when students perceived the instructors to hold a fixed view of intelligence, they experienced a lower sense of belonging, worse academic outcomes, higher concerns about being evaluated negatively, and higher impostor feelings and negative affect. Rattan and colleagues (2018) found that students who perceive STEM faculty to believe that most students have 'high scientific aptitude' report a higher sense of belonging and interest in STEM. "Fixed mindset professors are more likely to judge a student as having low ability based on a single test performance (Rattan et al., 2012) and to use unhelpful pedagogical practices, like encouraging students to drop difficult courses (e.g., "not everyone is meant to pursue a STEM career")" (Canning et al., 2019, p.1).

P-TOI is tangentially related to what is known as the Pygmalion effect (or expectancy effects)—the idea that teachers' expectations and beliefs about students have an impact on student outcomes (T. L. Good, 1987; Rosenthal, 1987, 1991; Rosenthal & Jacobson, 1968a, 1968b; Rosenthal & Rubin, 1978). However, the effects may be small and not additive (i.e., expectancy effects diminish over time; Rosenthal & Jacobson, 1968; West & Anderson, 1978). Nevertheless, small does not mean trivial, and small effects can still be meaningful, especially if they are moderated by contextual and individual differences (Jussim, 1990; Rosenthal, 1984, 1985). It has also been hypothesized that teachers' expectations in a classroom might reflect accurate judgments of students' abilities instead of expectations that become self-fulfilling (Jussim, 1986; Jussim & Harber, 2005), and this would confound any study of the effects of student perceptions. But the length and depth of interactions between K-12 teachers and their students might enable them to gauge their students' abilities accurately. College instructors, at least in the U.S., have limited interactions with students, which means their expectations are less likely to reflect accurate judgements. Indirect, non-specific cues about students' potential—how student questions are handled, how the course difficulty is framed, etc.—may be crucial pieces of information that students use to assess their potential for succeeding, especially in rigorous courses.

Insofar as instructors' mindsets may influence students' academic outcomes, there are two (non-mutually exclusive) ways it can manifest. Either the instructor's mindset affects their teaching philosophy and style (teaching quality as a mediator between mindset and student outcomes), or it can affect students' motivational processes (or some combination of the two). The only way to assess that is to gauge teachers' teaching quality and their mindsets simultaneously. How such perceptions form and how they influence student behavior important for academic success remains to be addressed.

Current Study

Although instruments measuring implicit theories of intelligence have been used extensively for a few decades, most are slight adaptations of the original six items (Dweck, 2000). Little attention has been paid to the psychometric properties of the existing items. Moreover, researchers have also argued that students may respond differently to items based on their personal definitions of intelligence, leading to measurement non-invariance (Limeri et al., 2020a, 2020b). Thus, developing valid and reliable measures in this domain is an important first step in assessing how these mindsets play out in classrooms. The current study aims to help address this issue by developing an instrument that measures students' perceptions of instructors' theories about intelligence using the *four building blocks* approach to measurement, delineated below (M. Wilson, 2005, 2023).

An instrument that measures the construct and points to specific behaviors that are linked to a malleable view of intelligence can be extremely beneficial for instructors, especially those who may otherwise not be sensitive to the cues about intelligence that they send students via pedagogical practices and interpersonal interactions. The primary hypothesis of the study is that Perceived Implicit Theories of Intelligence (P-TOI) is a unidimensional construct, with five hypothesized waypoints (levels), going from "High Fixed" at the lowest level to "High Growth" at the highest level. The construct is meant to mirror conceptualizations of individuals' theories about intelligence (from fixed to malleable); see the construct map below for a detailed description of each hypothesized level. P-TOI was hypothesized to have five levels based on practical and theoretical reasons. The traditional implicit theories items have a 6-point Likert response structure; attitudinal measures rarely yield scales that have more than four to five levels and generating more than five qualitative gradations (and labels) for P-TOI seemed unfeasible.

Use Case

In software and system development, a *use case* outlines the ideal functioning of a system as well as its potential applications. Here, we provide a use case for the P-TOI instrument, along with some cautionary notes. The instrument is meant to be used at the high school and undergraduate level by educators (and, given appropriate circumstances, administrators) interested in understanding students' sociopsychological experiences in classrooms. The instrument is expected to wield higher predictive power within STEM classes, as higher weight is given to intellectual ability for performing well in these classes (Bian et al., 2018; Rattan et al., 2018).

User

This instrument is meant for instructors/teachers or administrators interested in assessing students' perceptions of their instructor's beliefs about intelligence and ability and of the prevalence of classroom practices that cue perceptions of these beliefs.

Goal

The instrument should primarily be used to help improve pedagogical practices and not for accountability or evaluation. Care should be taken that instructors do not feel targeted or criticized about their teaching practices. Strict confidentiality of identifying student data should be maintained. Like with any feedback mechanism, approaching the use of the instrument as constructive (and not critical) is crucial.

System

The instrument is meant for use within the context of a particular classroom. Since the items ask students to assess various aspects of the teaching practice, they should be administered after the students have experienced all facets of the classroom environment. The optimal timing will vary by context, but a few weeks of classroom instruction should provide ample opportunities for students to become familiar with teaching and classroom practices. The instrument should ideally not be administered right before or after a major assessment to prevent spillover from test anxiety or concerns about grades. The instrument can also be administered at the beginning and end of the course to track shifts in student perceptions over time.

Scope

The instrument should be used only to get a better view of students' subjective experience and to make improvements to teaching practice considering that view. The instrument has been designed for and validated in STEM classrooms and is thus most appropriate for use in STEM courses. Only slight adaptations would need to be made for use in non-STEM courses, but given that perceptions about intelligence are less salient in other disciplines (Bian et al., 2018; Meyer et al., 2015), we recommend a thorough assessment of the psychometric properties of the instrument in those cases to ensure that it works as expected in those settings.

Four Building Blocks

This study uses an instrument development framework called the four building blocks of measurement (M. Wilson, 2005, 2023), a scientific approach to measurement that emphasizes theory development, the use of empirical evidence to test hypotheses, and iteration based on the results. The primary hypothesis in psychometrics is the validity of an instrument for measuring a construct, and the four blocks—these are the construct map, items design, outcome space, and calibration model—aid in testing this hypothesis. The construct map is a graphical representation of the underlying continuum of the construct of interest, going from the lowest level at the bottom to the highest level at the top. On the continuum are qualitative levels ('waypoints') that span the entire range of the theorized construct. The next building block, the design of the items, establishes a causal link between the construct and the empirical responses to the items, much like a thermometer establishes the link between a target and the theoretical concept of temperature. The items elicit responses that allow us to measure the latent construct and to map a target's locations on the construct using the next building block, the outcome space.

The outcome space represents a procedure for classifying observations into a set of standard categories that cover the entire domain of possible responses. (For forced-choice items, these categories are pre-determined.) Finally, once the empirical data have been collected and analyzed, the fourth building block, the calibration or statistical model, affords the creation of a quantitative scale underlying the qualitative levels of the construct. After the quantitative analysis has been conducted, the results feed into the refinement of both the instrument and the scientific theory underlying the construct, creating an iterative cycle between theory and empirical investigation.

Instrument Design

Construct Definition

Perceived Implicit Theories of Intelligence (P-TOI) refers to students' perceptions about their instructors' beliefs about the malleability of intelligence and ability. It is conceptualized as a unidimensional construct, with five hypothesized waypoints (levels), going from high fixed (students perceive the instructor to believe that students "either have it or they don't") at the lowest level and high growth at the highest level (students perceive instructor to believe that students can significantly improve their abilities). The construct is meant to mirror conceptualizations of individuals' own theories about intelligence (from fixed to malleable); see Figure 8 and Table 8 for a detailed description of each hypothesized level.

Generating Waypoints

Waypoints (i.e., the ordered qualitative levels) were based on the theoretical content of the construct and were meant to mirror the conceptualization of the traditional growth mindset construct (see Figure 8 for details). Table 9 highlights the feelings (Affect), actions (Behavior), and judgements (Cognition) of a hypothetical student at each hypothesized level of the construct. In other words, it is a theoretical representation of how a typical respondent at that level of the construct would think, feel, and act. [1]

[1] Note that this is for illustrative purposes only and the instrument focuses primarily on judgements/perceptions.

Figure 8

Perceived Theories of Intelligence Construct Map (V1)

Respondent	Responses
High Growth: Respondent perceives the instructor to hold the mindset that all students are capable of significant growth; seeks feedback from instructor and sees it as constructive and given in good faith; perceives instructor to be invested in every student's learning.	**High Growth:** "Instructor challenges student thinking, expects everyone to do well, and offers helpful strategies." "I feel really comfortable asking questions in class or asking for help." "Instructor turns student mistakes into learning experiences and normalizes failure." "Instructor is sure people will learn and improve in the class."
Growth: Respondent perceives the instructor to hold the mindset that many but not all students are capable of growth; welcomes feedback and sees it as helpful; perceives instructor to be invested in some students' intellectual growth.	**Growth:** "Instructor thinks that some students won't succeed in class, but most can if they put in effort." "I don't mind asking for help." ""Instructor thinks people will improve in class."
Mixed: Respondent perceives the instructor to hold the mindset that some students are capable of marginal growth; sees formative (but not critical) feedback as helpful; does not seek feedback or help	**Mixed:** "Instructor gives helpful feedback." "I don't really like asking questions in class." "Instructor thinks only some people will improve in class."
Fixed: Respondent perceives the instructor to hold the mindset that only a few students are capable of marginal growth; does not like getting feedback; avoids seeking help; perceives the instructor to strongly favor those with higher baseline ability	**Fixed:** "Instructor does not give good feedback." "Instructor thinks not all students will be able to excel in the class." "I don't go to office hours or ask questions."
High Fixed: Respondent perceives the instructor to hold the mindset that students are either smart or they're not; avoids asking for help; ignores feedback; perceives critical feedback to be a personal attack on intelligence; perceives instructor to only care about select few with higher baseline ability.	**High Fixed:** "Instructor encourages students to drop the course if they're struggling." "Instructor only focuses on the "smart" students." "I'm scared to ask for help." "I will look stupid if I ask questions in classroom."

Table 8

Affective, Behavioral, and Cognitive Experience of Hypothetical Student

High Growth (HGM)	Feelings (A)	Respondent feels very comfortable asking for help. Feels supported. Does not fear asking questions and is not worried about negative judgement. Feels motivated to do well.
	Actions (B)	Seeks feedback from instructor, asks questions in class. Goes to office hours when needed.
	Judgements (C)	Respondents perceive the instructor to hold the mindset that all students are capable of significant growth; sees feedback as constructive and given in good faith; perceives instructors to be invested in every students' growth.
Growth (GM)	Feelings (A)	Respondent feels supported and comfortable asking for help when needed. Does not feel stupid when asking questions and feels somewhat motivated to do well.
	Actions (B)	Welcomes feedback.
	Judgements (C)	Perceives the instructor to hold the mindset that many but not all students are capable of growth; sees it as helpful; perceives instructor to be invested in some students' intellectual growth.
Mixed (MM)	Feelings (A)	Respondent may feel comfortable asking for help in the beginning but that changes as the course progresses.
	Actions (B)	Respondent does not seek feedback or help.
	Judgements (C)	perceives the instructor to hold the mindset that some students are capable of marginal growth; sees formative (but not critical) feedback as helpful;
Fixed (FM)	Feelings (A)	Does not appreciate feedback; scared of being judged/asking questions.
	Actions (B)	Avoids seeking help
	Judgements (C)	Perceives the instructor to hold the mindset that most students' ability cannot change; perceives the instructor to strongly favor those with higher baseline ability
High Fixed (HFM)	Feelings (A)	Actively dislikes getting feedback; worried about being judged and afraid of asking questions
	Actions (B)	Avoids asking for help; ignores feedback;
	Judgements (C)	Perceives the instructor to hold the mindset that students are either smart or they're not; perceives critical feedback to be a personal attack on intelligence; perceives instructor to only care about select few with higher baseline ability.

Generating Items

Since the construct is a meta-perception (perception of another's perception about intelligence), the instrument aims to capture impressions or actions of someone who holds a malleable/entity view. The items were generated by studying instruments currently in use, reviewing the literature, and identifying behaviors and pedagogical practices that would convey instructors' theories about the malleability of intelligence (Aronson et al., 2022; Kroeper et al., 2022; Muenks et al., 2020; Rattan et al., 2012, 2018; Yeager et al., 2022). As the term 'intelligence' carries with it some baggage (Ritchie, 2015), it was not mentioned explicitly, and the relatively neutral term 'ability' was used instead. For example, if the instructor encouraged students, implicitly or explicitly, to show improvements in their work—by allowing them to resubmit assignments, giving constructive feedback, etc.—that would be an indication that the instructor emphasizes growth and improvement over assessing students' ability level.

Items Design

For validation purposes, the 13 items were recycled in four separate item bundles, each bundle asking the student to respond with a particular instructor/teacher in mind: one they considered effective at teaching (LI, for Liked as Instructor), one they considered ineffective at teaching (DI for Disliked as Instructor), one they liked as a person (LP), and one they disliked as a person (DP). Thus, students answered all items four times (52 total responses). The purpose was to disentangle students' general impression of their instructors from their implicit theories about intelligence. In hindsight, the item bundles may not allow us to clearly discriminate between the two; however, we do observe patterns that confirm certain expectations. We expected the presence of "construct-irrelevant variance" (AERA, APA, & NCME, 2014) from a halo effect (Nisbett & Wilson, 1977) such that instructors who were likable or effective would be rated as having a more malleable view of intelligence, which appears to be the case. The four response patterns also shed light on whether the instrument is measuring a perception of a target or an individual difference among respondents. If the items were capturing an individual tendency to rate all instructors positively vs. negatively, we should see parallel lines in Figure 9, which shows every student's mean score on all 52 items and the four item bundles. Although there are several parallel lines, there are many intersecting lines as well. In fact, the exaggerated chevron pattern indicates that those who gave more positive ratings to instructors they liked (LI, LP) rated the instructors they disliked (DI, DP) more negatively. Thus, the instrument does not appear to simply be capturing a tendency to give higher vs. lower ratings.

While responding to the survey, students were asked to indicate whether they could think of an instructor they considered effective, ineffective, likable, and unlikable, and if they answered in the negative, the survey skipped to the next section. If their answer was affirmative, they were asked to think about that instructor as they answered the item bundle that followed. Students could have considered the same instructor for more than one bundle, and the survey asked them to indicate if that was the case.

The item bundles began with the following stem for college students (See Table 9 for the complete instrument):

> "For this section, please think of a STEM instructor from one of your courses whom you [LIKED/DID NOT LIKE] AS AN INSTRUCTOR. That is, an instructor whom you considered to be <u>[effective/ineffective] at teaching</u>."

"For this next section, please think of a STEM instructor from one of your courses whom you [LIKED/DID NOT LIKE] AS A PERSON. That is, an instructor whom you [liked/disliked] for their personality, disposition, sense of humor, etc."

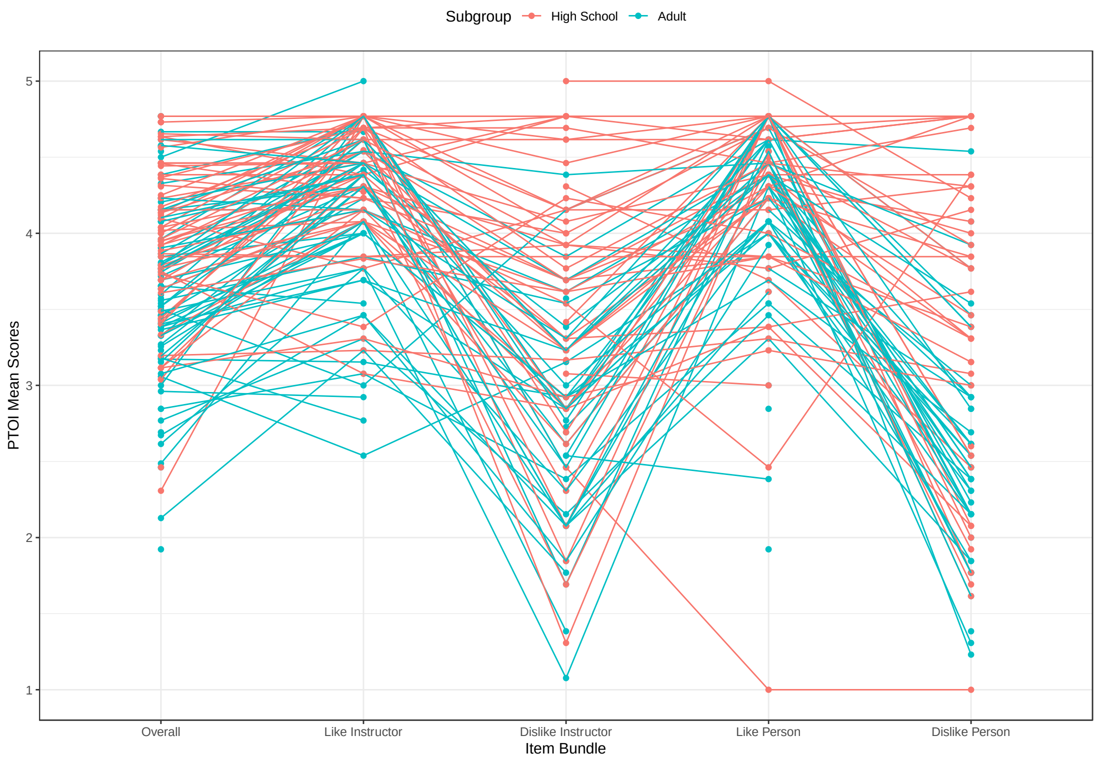
The item bundles for high school students began thus:

"For this section, we're going to ask you to think of a STEM teacher from one of your classes this year whom you [LIKE/ DO NOT LIKE] AS A TEACHER. That is, a teacher whom you consider to be [effective/ineffective] at teaching."

"For the next section, we're going to ask you to think of a STEM teacher from one of your classes this year whom you [LIKE/DISLIKE] AS A PERSON. That is, a teacher whom you [like/dislike] for their personality, disposition, sense of humor, etc."

Figure 9

Line Graph of Mean Scores in Study 2



Note. PTOI = Perceived Implicit Theories of Intelligence.

Outcome Space

The items assessed students' perceptions of their instructors' beliefs and theoretically relevant behaviors regarding student ability. Instead of the traditional Likert approach to item response design (e.g., *strongly agree* to *strongly disagree*), the instrument employed a Guttman approach to item response design (Guttman, 1944; M. Wilson et al., 2022), with each response option corresponding to a level on the construct. To illustrate, we can look at the first item on the instrument, Do Well (see Table 10); in the traditional approach, this item would be a statement, "My teacher thinks that all students can do well in this course," rated on a 5-point Likert scale. Using the Guttman approach, the item was instead separated into a stem ("My teacher thinks that…") and five options that aligned with each level of the construct. This ensured that the numerical values associated with the response options explicitly corresponded to levels on the construct map, and there was less room for subjective judgements about the meaning of a '4' on a 5-point scale. As there was a one-to-one matching between item responses and construct waypoints, students (implicitly) self-selected their level on the construct for each item; thus, the items were scored automatically.

Table 9

P-TOI Instrument

Item label	Item text	Waypoint
Do Well	My [instructor/teacher] thinks that…	
	…all students can do well in their course.	High Growth
	…most students can do well in their course.	Growth
	…many students can do well in their course.	Mixed
	…some students can do well in their course.	Fixed
	…only a few students can do well in their course.	High Fixed
Improve Abilities	My [instructor/teacher] thinks that…	
	…students in general are capable of significantly high growth.	High Growth
	…students in general are capable of high growth.	Growth
	…students in general are capable of some growth.	Mixed
	…students in general are capable of minimal growth.	Fixed
	…students in general are capable of no growth	High Fixed
Growth Possible (Quantitative)	My [instructor/teacher] thinks that…	
	…all students are capable of growth.	High Growth
	…most students are capable of growth.	Growth
	…many students are capable of growth.	Mixed
	…only a few students are capable of growth.	Fixed
	…no students are capable of growth.	High Fixed
Growth Possible (Qualitative)	My [instructor/teacher] thinks that…	
	…students in general are capable of significantly high growth.	High Growth
	…students in general are capable of high growth.	Growth
	…students in general are capable of some growth.	Mixed
	…students in general are capable of minimal growth.	Fixed
	…students in general are capable of no growth	High Fixed
Teaching Care	My [instructor/teacher]…	
	…cares about teaching all students.	High Growth
	...cares about teaching most of the students.	Growth
	…cares about teaching some students who are already doing well.	Mixed
	…cares about teaching a select few "bright" students.	Fixed
	…cares about teaching no students at all.	High Fixed
Actionable Feedback	My [instructor/teacher]…	
	…gives thoughtful, actionable feedback along with critical feedback.	High Growth
	…gives actionable feedback along with critical feedback.	Growth
	…sometimes gives actionable feedback along with critical feedback.	Mixed
	…only gives critical feedback with no actionable feedback.	Fixed
	…gives no feedback.	High Fixed
Growth Opportunities	My [instructor/teacher]…	
	…creates opportunities for students to show improvement (e.g. allowing students to resubmit assignments, dropping lowest scores, weighing later exams more heavily).	High Growth

Item label	Item text	Waypoint
	…does not create opportunities for students to show improvement.	High Fixed
Comfort Questions	In this [course/class], I feel…	
	…very comfortable asking questions.	High Growth
	…comfortable asking questions.	Growth
	…okay asking questions.	Mixed
	…uncomfortable asking questions.	Fixed
	…very uncomfortable asking questions.	High Fixed
Comfort Comments	In this [course/class], I feel…	
	…very comfortable speaking up/making comments.	High Growth
	…comfortable speaking up/making comments.	Growth
	…okay speaking up/making comments.	Mixed
	…uncomfortable speaking up/making comments.	Fixed
	…very uncomfortable speaking up/making comments.	High Fixed
Encourage	My [instructor/teacher]…	
	…strongly encourages and supports students.	High Growth
	…encourages and supports students.	Growth
	…encourages students.	Mixed
	…does not encourage students.	Fixed
	…demoralizes students.	High Fixed
Mistake Response	When students make mistakes in course, my [instructor/teacher]…	
	… responds kindly and nudges the student in the right direction.	High Growth
	…responds kindly and corrects students.	Growth
	… corrects students.	Mixed
	…moves onto other students without acknowledging.	Fixed
	…makes students feel stupid.	High Fixed
Normalize Failure	My [instructor/teacher]…	
	…thinks that struggle and failure are a part of STEM.	High Growth
	…thinks that struggle and failure are common in STEM.	Growth
	… thinks that struggle and failure may happen in STEM.	Mixed
	…thinks that failure and struggle mean low potential for STEM.	Fixed
	…thinks that failure and struggle mean one cannot succeed in STEM.	High Fixed
Teaching Improve	My [instructor/teacher] …	
	…asks for student feedback to make improvements to teaching practice.	High Growth
	…listens to students' feedback to make improvements to teaching practice.	Growth
	…listens to student feedback (when offered) but makes minimal effort to make improvements to teaching practice.	Mixed
	…does not listen to student feedback to make improvements to teaching practice.	Fixed
	…ignores student feedback and expends no effort to make improvements to teaching practice.	High Fixed

Calibration Model

We used item response theory (Thurstone, 1927, 1928) for modeling the data, and since the data were polytomous (5 ordered categories), a unidimensional partial credit model (Masters, 1982), an extension of the Rasch model, was used to calibrate the instrument. Data from the four item bundles were calibrated separately instead of calibrating all data with a single unidimensional or multi-dimensional model because each item bundle represents a different student-instructor pairing. Moreover, the 52 items together neither represent a single instrument nor four dimensions of a single construct, making an overall unidimensional or multi-dimensional model inappropriate for these data.

In the Rasch model, the probability of observing a response, X_i, for item *i* is modeled as a function of the respondent's level on the construct's continuum, denoted by θ, and how 'difficult' the item is to agree with, denoted by δ_i. In the simplest case (with binary outcomes), the probability of response $X_i = 1$ is given by the following equation:

$$\text{Probability}(X_i = 1|\theta, \delta_i) = \frac{e^{(\theta-\delta_i)}}{1 + e^{(\theta-\delta_i)}}$$

Generalizing the simple Rasch model to polytomous data involves thinking of the responses as a succession of binary outcomes. For example, if the ordered scores range from 0 to 4, like in our case, we can evaluate level 0 vs. level 1, level 1 vs. 2, level 2 vs. 3, and level 3 vs. 4, modeling the responses as dichotomous at each step (M. Wilson, 2005, 2023). The parameter estimates for each comparison are therefore called *step parameters* and reflect the probability of going from one 'step' (or level) on the construct to the next (Masters, 1982; Masters & Wright, 1981). In the case where the item responses range from 0, 1,…*m*, where *m* represents the steps, the probability of participant *p* scoring *x* on item *i* can be given by the following equation (representing the partial credit model):

$$p_{pix}(\theta) = \frac{exp\left[\sum_{j=0}^{x}(\theta_p - \delta_i - \tau_{ij})\right]}{\sum_{r=0}^{m_i}\left[exp\ \sum_{j=0}^{r}(\theta_p - \delta_i - \tau_{ij})\right]},\ x = 0, \ldots m_i,$$

$$where\ \sum_{j=0}^{0}(\theta_p - \delta_i - \tau_{ij}) = 0,$$

where θ represents the respondent *p*'s level on the construct, δ represents the difficulty of item *i*, and τ refers to the additional step parameter associated with category *j* of item *i* (Masters, 1982). The model was estimated with the ACER ConQuest software (Adams et al., 2012) using the marginal maximum likelihood estimator (Bock-Aitkin Quadrature with 80 nodes; Bock & Aitkin, 1981). The ability parameter θ_p was assumed to be normally distributed with $N(\mu_\theta, \sigma_\theta^2)$. Item parameters, δ_i and τ_{ij}, were assumed to be fixed. For model identification, the mean of the ability distribution and the sum of the step parameters was fixed to 0 for each item *i*.

The calibration model results in a 'scale' that can be used to compare respondents and items on a common metric (logits), which allows the researcher to graphically represent the relative positions of the items and respondents on the construct's continuum on a single visual display. This person-item map is also known as the Wright Map, named after psychometrician Ben Wright (M. Wilson, 2011; see Figures 11 and 12 for examples) and represents an empirical manifestation of the Construct Map.

Method

Participants

As a first step in the validation process, we collected data from both high school and college students to calibrate and investigate the instrument. Participants (N = 156) were recruited through word of mouth, an online data-collection platform, Prolific (https://www.prolific.co), and Character Lab Research Network (researcher-practitioner partnership at the University of Pennsylvania). The demographic composition of the entire sample is presented in Table 10. Most participants are from the United States, but adult participants recruited via Prolific represent 16 different countries (however, the sample from each country not including the US is small, ranging from 1–7).

Table 10

Sociodemographic Characteristics of Participants in Study 2

Baseline Characteristic	High School	Adults	Overall
Age	16.03 (1.30)	23.4 (5.63)	19.5 (5.40)
Gender			
Female	28 (43.1 %)	20 (22.0%)	48 (30.8%)
Male	34 (52.3 %)	36 (39.6 %)	70 (44.9%)
Missing	3 (4.6 %)	35 (38.5 %)	38 (24.4%)
Race/Ethnicity			
Asian/Asian-American	2 (3.1 %)	10 (11.0 %)	12 (7.7 %)
Black/African-American	14 (21.5%)	4 (4.4 %)	5 (3.2 %)
Hispanic/Latinx	24 (36.9 %)	9 (9.99 %)	33 (21.2 %)
White/European	15 (23.1 %)	33 (36.3 %)	48 (30.8 %)
Biracial/Mixed	1 (1.5 %)	4 (4.4 %)	5 (3.2 %)
Missing	9 (13.8 %)	35 (38.5 %)	44 (28.2 %)

Note. Information about gender was collected at the end of the survey and participants recruited through word of mouth were less likely to finish the survey, leading to a higher proportion of missing information.

Procedure

The data were collected during Fall 2020–Spring 2021 via Qualtrics (https://www.qualtrics.com). The Character Lab Research Network (CLRN) data were collected as part of a larger data collection effort that included a variety of studies designed by scientists affiliated with CLRN. CLRN simultaneously rolled out multiple independent studies, and students were randomized to one of the studies running in their school. This study was conducted on school computers during class time in participating schools over the course of a two-to-three-week testing window; data for this study were collected between Jan 1–Jan 3, 2021. The adult participants completed the survey between Dec 12–Dec 27, 2020. Adults were asked to think about an instructor from the current or a previous semester; the precise number of adults who were actively in college at the time of survey administration is unknown. No participants were compensated for the study.

Open Research Practice Statement

The study hypotheses and analysis plan were not preregistered. Post-hoc registration of the data collection procedure for the high school sample is available at https://osf.io/ahvb2. Data, code, and the entire survey can be found at https://researchbox.org/870.

Results

Item Fit

Table 11 presents the estimated item difficulty locations (in logits), and Table 12 shows the item fit results for each item bundle (calibrated separately). The same information is visualized in Figure 10; item difficulties in Panel A and Weighted *MNSQ* values in Panel B.

Looking at the estimated item difficulties presented in Table 11, we can see that in most cases, LI and LP estimates are quite close, as are the DI and DP estimates. The standard error associated with these estimates ranges from 0.12 to 0.33, which gives some perspective on the differences. Most values fall between .12–.18, with the exception of Growth Opportunity, which has standard errors ranging from .25–.33; this is to be expected given that it was the only dichotomous item in the set. LI and LP items are easier to agree with (lower on Panel A in Figure 10), and DI and DP are harder to agree with.

Internal structure at the item level was evaluated using a metric called the Mean Square (*MNSQ*) Fit Statistic (Masters & Wright, 1981), calculated by taking a ratio of two variances (expected and observed squared residuals for items), which allows us to assess how important an item misfit is. If the observed residuals are perfectly aligned with the expected residuals, the value of the *MNSQ* statistic is 1. If the value is greater than 1, the observed variance is greater than expected; conversely, if the value is less than 1, the observed variance is smaller than expected. According to M. Wilson (2005, 2023), items with *MNSQ* values greater than 1 should be attended to first as they are noisier. Although there are no standard criteria for evaluating a weighted MNSQ value, researchers have provided a general rule of thumb: 3⁄4 and 4⁄3 (Adams & Khoo, 1996). That is, items that have *MNSQ* values lower than .75 or higher than 1.33 can be thought of as misfitting. In addition, one can also look at the weighted *t*-statistic, which is calculated by transforming the weighted *MNSQ* values into a standard normal distribution (see Table 12), but this value can be significant for many items when the sample size is large (Wright & Masters, 1981). Therefore, the instrument developer is advised to look at both the *MNSQ* value and the *t*-statistic and make a context-specific judgement (M. Wilson, 2005, 2023).

The observed variance appears to align with the expected variance for most items, with *MNSQ* values ranging from .70 to 1.57. Values for only 7 out of the 52 administered items falling out of the .75–1.33 boundary (represented by gray dotted lines in Figure 10 Panel B). All four Encourage items fall lower than the .75 threshold, indicating that this item had lower variance than expected (i.e., it is more highly correlated with the rest of the items). The small set of items that have *MNSQ* values above 1.33 are slightly more concerning since these items are noisier than expected by the calibration model. This set includes Comfort Comments from the Like as Person item bundle and Normalize Failure from both Like as Instructor and Dislike as Person item bundles. The item Comfort Comments inquires whether students feel comfortable making comments or speaking up in class, and it is likely that this item is affected to a greater extent by individual differences (e.g., personality), cultural background, or classroom environment. The reason for the higher variance in responses to Normalize Failure remains opaque, especially since the two misfitting items are from item bundles relating to instructors students liked *and* disliked, making this result difficult to interpret.

Although items with *MNSQ* values outside the .75–1.33 boundary on the lower end are less worrisome, the Encourage item has lower *MNSQ* values for *all* four item bundles. This item asks whether the student perceives the instructor as encouraging and supportive (on the high end) or demoralizing (on the low end). This result indicates that the item was highly correlated with other items in the instrument. This can be confirmed by looking at the item-rest correlations for this item, which are higher than the rest and range from .82–.88. Qualitatively, this could mean that the instrument is capturing a latent construct that is more akin to *teacher supportiveness* than beliefs about intelligence. We explore this possibility further in the discussion.

Table 11

Item Difficulties and Standard Errors in Study 2

Item	Like as Instructor		Dislike as Instructor		Like as Person		Dislike as Person	
	δ	*SE*	δ	*SE*	δ	*SE*	δ	*SE*
Do Well	-2.18	0.16	-0.78	0.12	-3.12	0.17	-0.26	0.13
Improve Abilities	-2.79	0.18	-0.98	0.13	-3.31	0.18	-0.80	0.18
Growth Possible Quant)	-2.87	0.18	-1.61	0.13	-3.32	0.17	-1.85	0.15
Growth Possible (Qual)	-1.26	0.18	-1.33	0.15	-2.92	0.17	-1.52	0.15
Teaching Care	-2.41	0.21	-1.00	0.12	-3.03	0.18	-0.96	0.17
Actionable Feedback	-2.71	0.17	-0.04	0.12	-3.07	0.16	-0.18	0.14
Growth Opportunity	-2.58	0.31	-0.46	0.25	-2.90	0.33	-0.26	0.26
Comfort Questions	-2.09	0.15	0.14	0.13	-2.12	0.14	0.47	0.14
Comfort Comments	-1.91	0.18	0.22	0.13	-2.15	0.14	0.38	0.14
Encourage	-2.45	0.17	-0.27	0.14	-2.22	0.15	0.12	0.14
Mistake Response	-2.54	0.16	-0.48	0.13	-2.47	0.15	-0.16	0.14
Normalize Failure	-2.21	0.16	-0.88	0.13	-2.43	0.15	-1.01	0.14
Teaching Improve	-2.35	0.15	-0.08	0.13	-2.89	0.17	0.03	0.14
Mean Difficulty (*SD*)	-2.33 (0.43)		-0.58 (0.57)		- 2.77 (0.43)		-0.46 (0.71)	

Note. δ = Item Difficulties.

Table 12

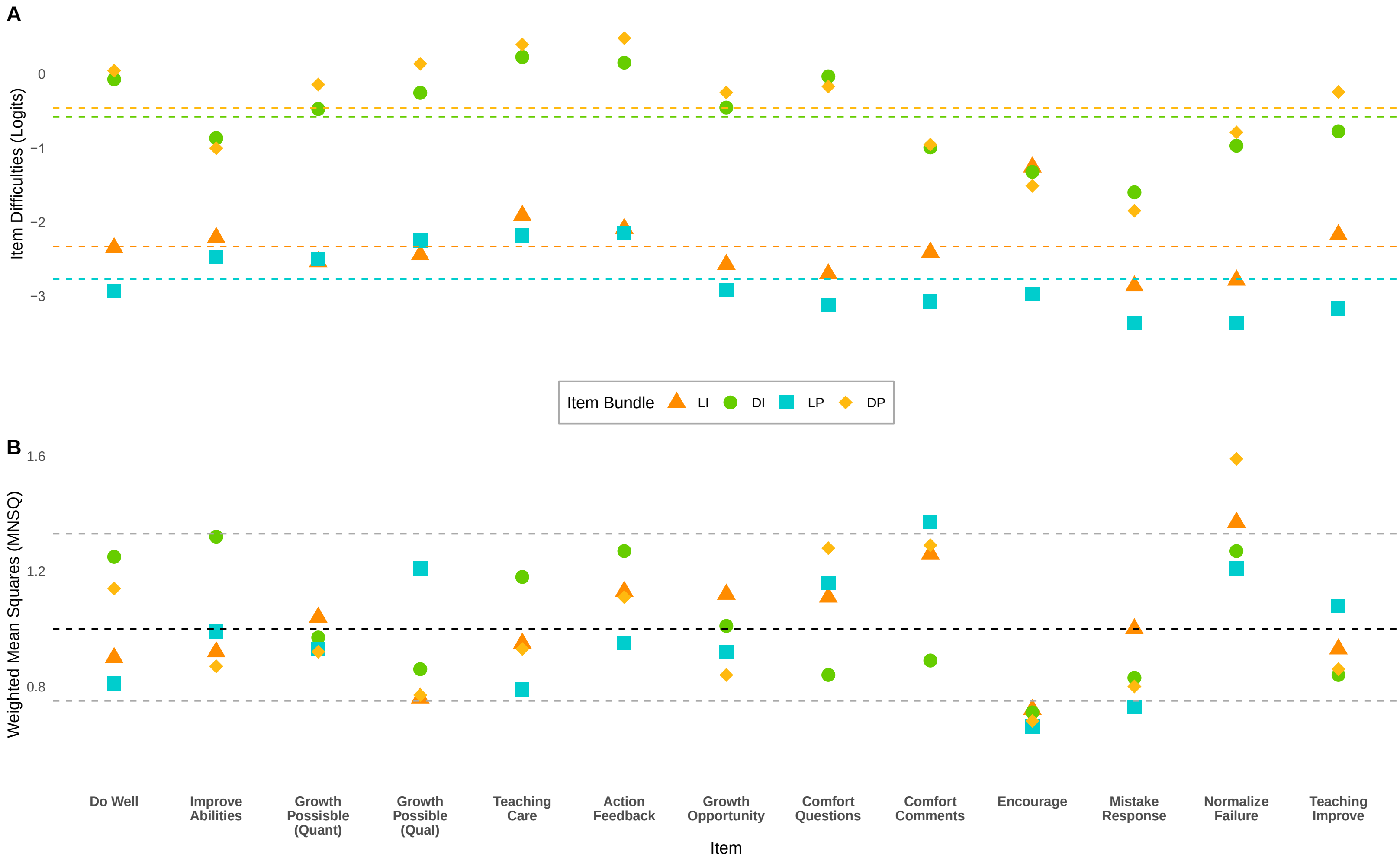
Mean Square Values and Item Fit in Study 2

Item	Like as Instructor				Dislike as Instructor				Like as Person				Dislike as Person			
	MNSQ	CI	*t*	*p*	*MNSQ*	CI	*t*	*p*	*MNSQ*	CI	*t*	*p*	*MNSQ*	CI	*t*	*p*
Do Well	0.91	(0.70, 1.30)	-0.6	.55	1.24	(0.71, 1.29)	1.5	.14	0.82	(0.67, 1.33)	-1.1	.27	1.15	(0.68, 1.32)	0.9	.37
Improve Abilities	0.91	(0.68, 1.32)	-0.5	.62	1.31	(0.71, 1.29)	1.9	.06	0.98	(0.65, 1.35)	-0.0	.99	0.86	(0.70, 1.30)	-0.9	.37
Growth Possible (Quantitative)	1.05	(0.65, 1.35)	0.3	.76	0.95	(0.71, 1.29)	-0.3	.76	0.94	(0.67, 1.33)	-0.3	.76	0.92	(0.68, 1.32)	-0.5	.62
Growth Possible (Qualitative)	0.75	(0.75, 1.25)	-2.1	.04	0.87	(0.71, 1.29)	-0.9	.37	1.19	(0.70, 1.30)	1.2	.23	0.80	(0.70, 1.30)	-1.4	.16
Teaching Care	0.94	(0.65, 1.35)	-0.3	.76	1.12	(0.70, 1.30)	0.8	.43	0.78	(0.54, 1.46)	-0.9	.37	0.89	(0.69, 1.31)	-0.7	.49
Actionable Feedback	1.12	(0.65, 1.35)	0.7	.49	1.25	(0.72, 1.28)	1.7	.09	0.96	(0.70, 1.30)	-0.2	.84	1.15	(0.70, 1.30)	1.0	.32
Growth Opportunities	1.12	(0.63, 1.37)	0.7	.49	0.99	(0.81, 1.19)	-0.0	.99	0.89	(0.59, 1.41)	-0.5	.62	0.85	(0.76, 1.24)	-1.3	.20
Comfort Questions	1.14	(0.71, 1.29)	0.9	.37	0.85	(0.72, 1.28)	-1.0	.32	1.11	(0.69, 1.31)	0.7	.49	1.27	(0.71, 1.29)	1.7	.09
Comfort Comments	1.26	(0.72, 1.28)	1.7	.09	0.91	(0.72, 1.28)	-0.6	.55	1.33	(0.73, 1.27)	2.2	.03	1.33	(0.69, 1.31)	1.9	.06
Encourage	**0.70**	**(0.73, 1.27)**	**-2.4**	**.02**	**0.72**	**(0.71, 1.29)**	**-2.1**	**.04**	**0.62**	**(0.67, 1.33)**	**-2.6**	**.01**	**0.69**	**(0.71, 1.29)**	**-2.3**	**.02**
Mistake Response	0.98	(0.68, 1.32)	-0.1	.92	0.83	(0.70, 1.30)	-1.2	.23	0.69	(0.67, 1.33)	-2.0	.05	0.81	(0.71, 1.29)	-1.3	.20
Normalize Failure	**1.45**	**(0.71, 1.29)**	**2.7**	**.01**	1.26	(0.70, 1.30)	1.6	.11	1.22	(0.72, 1.28)	1.5	.14	**1.57**	**(0.69, 1.31)**	**3.2**	**.002**
Teaching Improve	0.95	(0.71, 1.29)	-0.3	.76	0.84	(0.72, 1.28)	-1.1	.27	1.06	(0.69, 1.31)	0.4	.69	0.85	(0.70, 1.30)	-1.0	.32

Note. *MNSQ* = Mean Square. Items with *MNSQ* estimates outside the 95% confidence interval as well as the .75–1.33 range are in bold.

Figure 10

Item Difficulties and Mean Square Values in Study 2



Note. LI = Liked as Instructor; DI = Disliked as Instructor; LP = Liked as Person; DP = Disliked as Person. Colored dotted lines in Panel A represent average difficulty for each item bundle. Black dotted line in Panel B represents the expected value of the *MNSQ* statistic (1.00), and gray dotted lines indicate the .75–1.33 boundary.

Reliability

Reliability is an index used to assess the instrument's precision and the level of measurement error. As a unit-free index, reliability provides a useful tool for quantifying the uncertainty of measurement of individuals associated with the measurement process. Table 14 shows the *EAP* (Expected A Posteriori) reliability for each model, which represents the reliability of the respondents' scaled *EAP* scores. The table also presents the number of parameters estimated in each model along with the traditional index of internal reliability, Cronbach's α (Cronbach, 1951). All four item bundles are acceptably reliable, and Dislike as Person data shows slightly higher reliability.

Table 13

Reliabilities for Four Item Bundles in Study 2

Item Bundle	All Data				Complete Data			
	n	*EAP* Reliability	Cronbach's α	Parameters	*n*	*EAP* Reliability	Cronbach's α	Parameters
Like as Instructor	115	.86	.90	41	59	0.88	0.92	39
Dislike as Instructor	94	.94	.94	50	58	0.92	0.93	50
Like as Person	121	.87	.93	50	59	0.93	0.94	50
Dislike as Person	99	.95	.96	50	56	1.00	0.97	50

Note. *EAP* = Expected A Posteriori.

Validity

One of the most important goals of the measurement process is to develop an instrument that accurately measures what it purports to measure. The trustworthiness of an instrument can be considered to have two components (Mari et al., 2023): *objectivity* (does the instrument convey information solely about the construct of interest) and *intersubjectivity* (does the instrument convey the same information for different subjects/contexts). In the current study, the argument for intersubjectivity can be bolstered in two ways. First, the data were collected from both high school and adult participants, and second, the sample includes participants from 16 different countries (although, as mentioned earlier, the sample sizes for each country are small). We evaluate this argument with Differential Item Functioning (DIF) analysis in a later section. Arguments for objectivity are presented below.

Content Validity

Evidence Based on Test Content

Primary evidence based on test content comes from the development of the construct map (see Figure 8 and Table 8), which was based on the theoretical and conceptual tenets of the implicit theories of intelligence construct.

Evidence Based on Response Processes

At the end of each sub-section (for each instructor) in the instrument, students were asked to report their thought processes while answering the items. The prompt went as follows: "Please write down what you thought about as you answered these questions. Write down specific instances, examples, etc., if they came to mind. Please be as thorough as possible." The goal was to assess whether student reports were aligned with the intended construct. We received 81 relevant responses for instructors considered good at teaching, 63 for teachers considered not good at teaching, 87 for instructors students liked, and 66 for ones students did not like. Table 15 presents a selected list of student responses in each category. (For a list of all open-ended responses, please refer to the Supplementary Material.)

In their open responses, students considered various teaching practices regarding how instructors answered questions, created classroom structures, and interacted with students. (We should, however, note that the content of the instrument likely primed students to respond in a way that aligned with the content of the instrument.) Although by no means conclusive, informal qualitative analysis of the responses revealed that the instructors most likely to get responses we would expect for those perceived to hold a malleable view of intelligence were instructors students liked as people; students reported that likable instructors provided detailed feedback, were less intimidating, and relayed the belief that all students can perform well.

Effective instructors were unsurprisingly reported to employ effective teaching practices (provided good explanations, made complex concepts easier, ensured that students understood the content, had an engaging teaching style, gave personalized feedback, and displayed enthusiasm for subject); had a supportive and caring attitude (were approachable and easy to talk to, offered support outside of class, gave individual attention, and were responsive to student feedback); were fair and accommodating; promoted active engagement and learning; and helped students reflect on their study habits.

Instructors considered ineffective at teaching were deemed knowledgeable but unable to communicate or answer questions effectively; delivered lectures in an uninteresting or unorganized manner; showed a negative attitude (humiliated or condescended to students, dismissed student concerns, compared students to peers); favored some students over others; did not provide adequate support or guidance; set unclear expectations; were unavailable outside of class; and had unfavorable course policies (did not provide enough practice or feedback, did not prepare students enough for exams).

Instructors whom students considered likable had results similar to those considered good at teaching and, there was some overlap between the two (multiple students responded with the same instructor in mind). Students mentioned that these instructors made the class more enjoyable and engaging, used humor and personal stories, made students feel comfortable participating and asking questions without fear of judgement, showed genuine concern for students' well-being, fostered a sense of community, set clear expectations, and encouraged personal growth. They were also deemed to engage in effective teaching practices, as mentioned above, were open to student feedback and input, gave personalized feedback, and encouraged better study strategies. They expressed confidence in students' abilities to succeed and created environments that were learning and not performance focused. Students also reported finding them approachable and feeling a personal connection with these instructors.

Instructors students disliked were reported as creating negative classroom environments (belittling, dismissing, or mocking students), not providing adequate feedback or explaining mistakes, being ineffective at teaching (teaching methods not suitable for the content, failing to provide sufficient explanation), lacking in enthusiasm and passion (negative attitude toward teaching), unapproachable and insensitive (mean jokes or singled out students), not receptive to questions or requests for clarification, created an atmosphere of competitiveness. Students reported difficulty communicating with these instructors and noted that they gave vague and unhelpful responses.

It is important to note that a subset of students indicated difficulty in assessing their instructors' personal beliefs: "Many of these questions required me to guess about what thoughts my instructor had, which is very difficult and I found myself trying to deduct [sic] what the instructor would have thought from various actions that might not actually reflect what they believe." This concern is valid, but as we are interested primarily in students' subjective judgement, which need not be accurate, we do not share this concern to the same extent. There were also many high school students who could not think of teachers they disliked or considered ineffective. There are two possible reasons for this. Students might be reluctant to report negative evaluations of their teachers in case their responses were made available to teachers. Alternatively, it is also possible that the schools participating in the Character Lab Research Network (CLRN) are not representative of typical high schools.

The open-ended responses confirmed our suspicion of overlap between liking instructors and perceiving them to believe in the intellectual potential of all students. The current study does not allow us to uncover the causal direction of this relationship and is something future studies (preferably in the lab), with the aid of the P-TOI instrument, could examine.

Table 14

Select Responses to Open-Ended Items in Study 2

Liked as Instructor
"For most questions about my perception of the instructor's intentions or beliefs about students I tried to remember how my instructor responded to student questions. My reasoning for giving positive answers in most of these instances was times the instructor went from student to student to make sure they all understood the material, or created small groups within the classroom to help students learn better with their peers. So I tried to think of actual behaviors. I did feel biased though because I did well in the class, it is hard to imagine how someone who did poorly would perceive the instructor."
"I am thinking about how this professor told stories and jokes about their own mathematical growth and experiences. This made it easier to connect to the material and to feel confident about our own abilities (since we got to hear about his mistakes)."
"Actively tries to teach complex mathematical topics in a way that starts easy to understand, so that students were not initially deterred by some arcane symbols in a definition that is in reality quite simple. Always made sure to have students at least get part of the way to the answer by themselves in office hours to give them the skills to attempt similar problems on their own"
Disliked as Instructor
"The teacher was nice and kind and I did like her, however she did not teach very well and almost seemed like she had trouble understanding difficult concepts herself. Overall I thought about how I felt she was untrained in handling some situations where students were failing. She did offer extra

	help but unfortunately many kids did not make an effort to go because of the way they felt towards this teacher would not help them improve.”
	‘I remember her being a condescending individual that would use public humiliation as punishment and would make students cry. A friend of mine ran out of her room before in tears quite a few times because of his [sic] this professor treated her.”
Liked as Person	
	“This professor was AMAZING! It was so easy to choose them as a professor I liked as a person. For every single question asked I could think of a behavior of the professor (them saying they believe all students can do well, providing ear plugs and fidget toys during exams, giving detailed feedback on every assignment, etc.)”
	“This class was so much fun. The professor always made the class enjoyable and interesting. This was one of the few classes in which I was not afraid to raise my hand because I felt like I wasn't going to be judged for saying something wrong.”
	“This lecturer (and the previous, when I recollect) both talked a fair bit less about failure etc. They both just really liked maths, and it made it easy for others to like maths too. "Success" or "failure" were in some deep sense less salient in the class, because they're not really maths, so they're not really what the teacher cared about. And I don't mean that in the trite sense of "oh grades don't matter, LEARNING matters" - if you've met a mathematician, you know what I mean; they just genuinely only care about the maths and want you to care about it too. That took the pressure off and helped you to understand the maths, since you inherited some of their enthusiasm for maths purely by being in the class.”
	“There was constant feedback cycles and a willingness to listen to student complaints and suggestions. There was also a willingness to improve beyond expectation, and the professor created and fostered a collaborative environment that supported student growth by being open to all questions and setting clear guidelines for what the outcomes of the class in terms of grades would look like”
Disliked as Person	
	“While answering these questions, I got flashbacks of the class and how much I hated it. I always felt out of place in the course and never felt safe to ask questions. I always felt dumb and the instructor intimidated me. Whenever I asked for them I ended up feeling dumber and more confused but I was always scared to say so, so I would just say "Okay I get it, that makes sense” when in reality I was incredibly confused.”
	“The lecturer literally said on the first lecture that most students were likely not capable of understanding his course and should unenroll. To his credit, he was quite patient when we made mistakes, but he never quite dropped the whole "I'm only here for the future Nobel laureates" vibe.
	“With this professor I always felt stupid during assignment because they always made them more difficult than necessary and when they explained the mistakes they always made me feel stupid and made mean jokes about students in needs"

Along with the open-ended response process items, we also asked participants for direct feedback about the instrument. We included two multiple-choice items (10-point Likert-scale) that were followed by an open-response text box to evaluate how engaging participants found the survey and how relevant it was to their life. We also included two open-ended items that inquired whether students’ peers would find anything confusing and whether the survey language could be adapted to improve readability and comprehension. These items were included in the survey but not administered to the high school sample, and thus, we only have responses from the adult sample.

Participants (n = 52) reported finding the survey moderately engaging (M = 6.83, SD = 2.03) and equally relevant to their lives (M = 6.37, SD = 2.92). Although participants found the survey interesting, many reported finding it lengthy and repetitive, which is understandable given that they answered the same items four times. Some found the items to be thought-provoking and enjoyed reflecting on their experiences ("I thought it was fun to reflect on my years of schooling and the professors that stood out to me most. I found it easy to think of people I completely loved vs did not like."). Some reported finding the open-ended items more engaging than the multiple-choice items, which one person thought were "horrid" (stated without reason). Participants who reported not finding the survey relevant to their lives listed reasons such as not taking many STEM courses or not interacting with their instructors. A few reported the items as irrelevant to a non-U.S. context, and although not relevant to the instrument itself, many stated that the GPA item could be confusing for their peers. Participants did not have trouble understanding the instrument, and there was no actionable feedback for improving item text.

The primary respondent criticism of the instrument was the repetitiveness of the survey and the cognitive overload from having to remember which instructor was being referenced in each item bundle. However, since those features were present only for validation purposes, the criticism does not apply to the main instrument. Absent these features, it is possible that participants would have provided additional, substantive feedback; we hoped to capture some of this feedback by asking targeted questions about the instrument (ease of comprehension, relevance to life, etc.) later in the survey. Based on the feedback, it is currently recommended that the survey only be used in the United States as it may not apply to individuals outside the American context.

Wright Maps

Wright maps are visual representations of the item and person fit and allow us to visualize the distribution of respondents (on the left) and the items (on the right) using the same scale (logits), which represents the construct's continuum. This visualization makes it easy to see the relative position of respondents and items on the underlying construct. The item side displays Rasch-Thurstone thresholds, and there are four thresholds for five levels of the construct. The thresholds signify where a respondent's location (in logits) must be to have precisely a 50% probability of being at the level above or below (Linacre, 2010; M. Wilson, 2005, 2023).

Below we present two Wright maps, one for the item bundle that had lower reliability than the rest and the poorest evidence for banding—Liked as Instructor—and one that represents the highest reliability and some evidence of banding—Disliked as Instructor. *Banding* refers to whether there is a clear demarcation between the item parameters when going from one level to the next (see section Internal Structure Evidence). As might be expected, the Wright map for Liked as Person was similar to Liked as Instructor, and the one for Disliked as Person was similar to Disliked as Instructor and are thus not shown here (but are available in Appendix C). Disliked and Instructor and Disliked as Person likely show cleaner bands because of higher variation in scores (see Figure 9).

The person distribution (on the left) for Liked as Instructor (Figure 11) shows a slight left-skew, and more students fall at the high end of the distribution. This is to be expected as more respondents endorsed the higher end of the scale for this item bundle. The step parameters (on the right) for the two lower levels of the construct show considerable overlap, and in fact, some of the estimates for the second step parameters (the step from Fixed Mindset, FM, to Mixed Mindset, MM) are lower than the first step. In the context of hypothesis testing or student assessment, an argument could be made for combining the two lower levels. However, since these data were collected for instrument validation, and given that the other item bundles do show evidence for five levels, we do not presently make that recommendation. The Disliked as Instructor Wright Map (Figure 12) shows a more normal person distribution on the left and more clear evidence of banding on the right, with less overlap between step parameters. Overall, responses for instructors students did not like have higher variance (also evident in Figure 9), evenly spaced cutoff points (represented by gray lines in Figure 12), and better fit to the hypothesized structure of the construct.

Internal Structure Evidence

For one item bundle, Like as Instructor, there is evidence for fewer levels than hypothesized. Some items have three instead of four steps since no participant endorsed the lowest level for the instructor they considered effective. Consequently, data from the Like as Instructor item bundle shows the poorest fit to the hypothesized structure of the construct. This is also reflected in fewer estimated parameters (9 to be exact; see Table 13) for this item bundle. Although it is not the case for instructors students liked as people, some thresholds are close enough to warrant combining the waypoints for those data. The steps have a higher dispersion for instructors that the students did not think were good instructors and ones they did not like, which is to be expected given the high variance of the raw scores. The items are generally harder to endorse for these instructors as well and relatively easier to endorse for instructors students liked.

Banding

A Wright Map plots the relative difficulties of items along the levels (waypoints) of the construct, thus allowing us to investigate the hypothesized levels as “bands” in the Wright map. This investigation is based on a visual inspection of how the item-step difficulties are distributed along the continuum and a qualitative judgement by the investigator on where the bands should be placed to denote graduation from one level to the next. Wright maps in Figures 11 and 12 have been fitted with cutoff points (based on subjective judgement) to reflect bands that represent the different levels. There is little evidence for banding in the LI data, which, as mentioned above, reflected the poorest fit to the model. The DP data shows the best evidence for banding, along with the highest reliabilities, indicating a better fit to data and the hypothesized construct map.

Figure 11

Liked as Instructor Wright Map

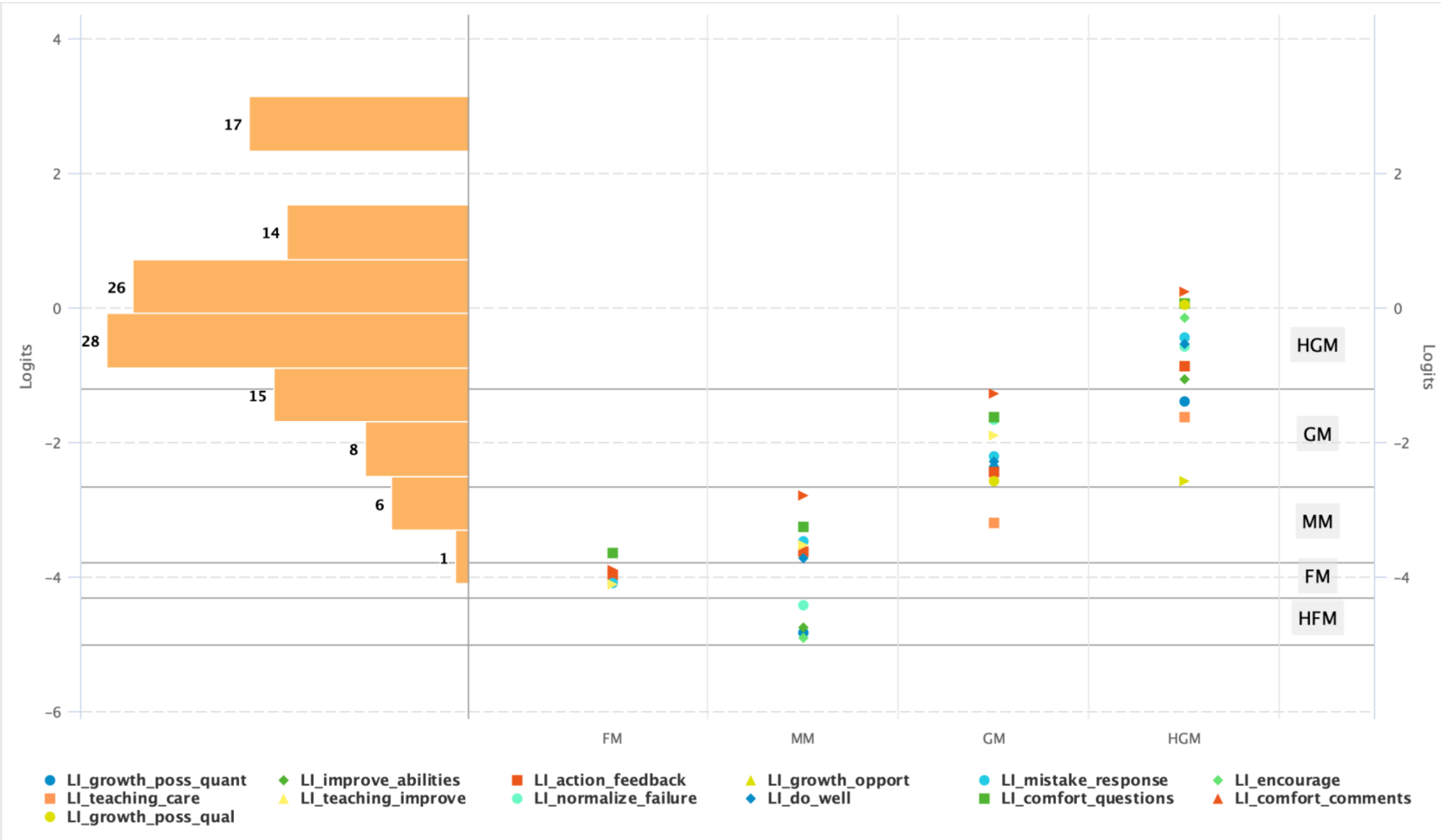


Figure 12

Disliked as Instructor Wright Map

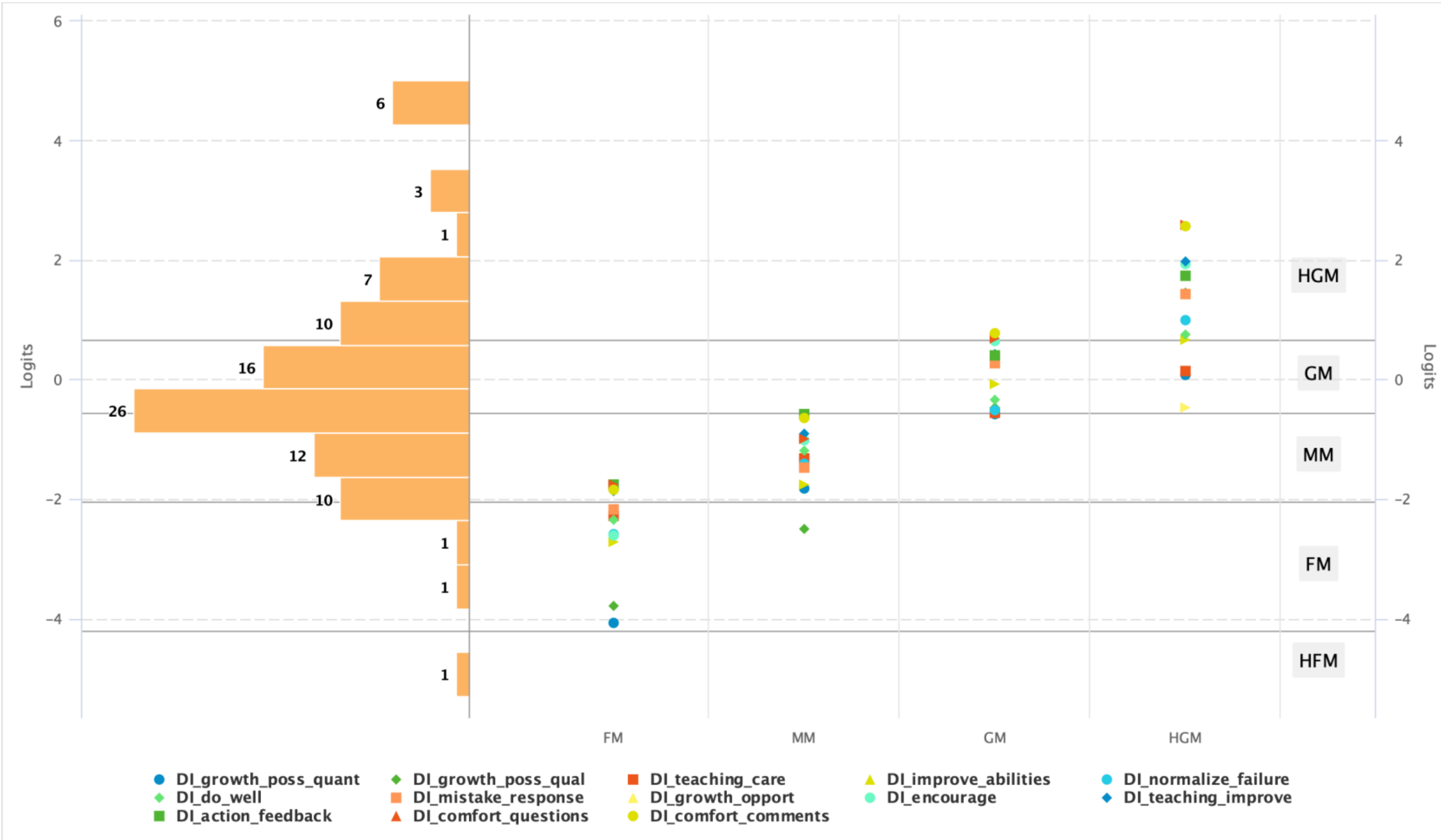


Table 15

Instruments/Items Included in Study 2

Instrument	Description	M	SD	Cronbach's α
Help-seeking (MSLQ)	5-item subscale from the Learning Strategies section of The Motivated Strategies for Learning Questionnaire (MSLQ) that measure students' tendency to manage support and seek help with coursework. *Source*: Pintrich & DeGroot (1990) *Response options*: 7-point Likert scale from 1 = *not at all true of me* to 7 = *very true of me* *Example item*: "I ask the instructor to clarify concepts I don't understand well."	4.14	1.08	.53
Help-seeking as Threatening	3 items that measure students' concerns about being assessed negatively by instructors and peers for seeking help. Followed by an open-ended item that asked students for reasons why they might not seek help even when they need it. *Source*: Skaalvik & Skaalvik (2005) *Response options*: 5-point Likert scale from 1 = *false* to 5 = *true* *Example item*: "I worry that other students may think that I am stupid if I ask for help."	2.10	1.02	.85
Evaluative Concern	5-item instrument that asked students whether they worried about their intelligence or abilities being judged negatively by their instructor. *Source*: Muenks et al. (2020) *Response options*: 7-point Likert scale from 1 = *not at all* to 7 = *a great deal* *Example item*: "On a typical day in this instructor's course... how much would you worry that the instructor might think that you are a slow learner?"	2.98	1.58	.90
Perceived Faculty Growth Mindset	4-item instrument that measures students' perceptions of their instructors' beliefs about the malleability of intelligence *Source*: Muenks et al. (2020) *Response options*: 6-point Likert scale from 1 = *strongly agree* to 6 = *strongly disagree* *Example item*: "In general, most professors at my institution seem to believe that students have a certain amount of intelligence, and they really can't do much to change it."	4.35	1.13	.91
Growth Mindset	3-item instrument that assesses beliefs about the malleability of intelligence. Followed by an open-ended item asking students to provide their personal definition of intelligence. *Source*: Dweck (2000); Yeager et al. (2016) *Response options*: 6-point Likert scale from 1 = *strongly agree* to 6 = *strongly disagree* *Example item*: "Your intelligence is something about you that you can't change very much."	4.07	1.13	.89
Academic Self-Efficacy	11-item instrument measuring students' level of confidence about successfully accomplishing academic tasks *Source*: Zimmerman et al. (1992) *Response options*: 11-point slider scale from 0 to 100 (10-point intervals) from 0 = *no confidence at all* to 100 = *complete confidence* *Example item*: "How much confidence do you have that you can successfully: finish assignments by deadlines?"	6.11	1.78	.87

Note. MSLQ = Motivated Strategies Learning Questionnaire

Evidence Related to External Variables

Convergent Validity

To test convergent validity, we collected ratings on the items used in previous research on perceived faculty growth mindset (Muenks et al., 2020) from a subset of the college student sample (n = 52). These items measure the same construct as the one of interest here and mirror the original Implicit Theories of Intelligence (ITOI) items (Dweck, 1999). We also measured *evaluative concern* (Muenks et al., 2020), *perceptions of help-seeking as threatening* (Skaalvik & Skaalvik, 2005), *self-efficacy* (Zimmerman et al., 1992), and self-reported *help-seeking behavior* (Pintrich & DeGroot, 1990).[2] All administered measures are listed in Table 16, and all items can be found in Supplementary Material.

The correlations between the four faculty growth mindset items and the P-TOI instrument range from -.10 to .57, depending on the target. Correlations are positive and significant for instructors students liked: Liked as Instructor, $r(49) = .57$, $p <.001$, 95% CI [.34, .73]; Liked as Person, $r(49) = .39$, $p = .005$, 95% CI [.13, .60]. Correlations are low or negative and not significantly different from zero for instructors they disliked: Disliked as Instructor, $r(34) = -.10$, $p = .57$, 95% CI [-.41, .24]; Disliked as Person, $r(38) = .06$, $p = .69$, 95% CI [-.25, .37]. When we limit the data to participants who answered all four item bundles (n = 26), the correlation with faculty growth mindset items gets stronger for Liked as Instructor $r(24) = .61$, $p = .001$, 95% CI [.29, .81], Disliked as Instructor, $r(24) = -.23$, $p = .26$, 95% CI [-.57, .17], and Dislike as Person, $r(24) = .13$, p = .51, 95% CI [-.27, .50]. The correlation gets weaker and is too small to reach the significance threshold given the sample size for Liked as Person, $r(24) = .33$, $p = .10$, 95% CI [-.07, .63]. Correlation matrices derived from the full and subsetted datasets are presented in Table 16.

Table 16

Correlations Among Faculty Growth Mindset and P-TOI Item Bundles

	Faculty GM	Like as Instructor	Dislike as Instructor	Like as Person	Dislike as Person
Faculty GM		0.61***	-0.23	0.33	0.13
Like as Instructor	0.57***		0.21	0.54***	0.19
Dislike as Instructor	-0.10	0.26*		0.28*	0.51***
Like as Person	0.39**	0.55***	0.31**		0.27*
Dislike as Person	0.06	0.21	0.53***	0.31**	

Note. Faculty GM = Faculty Growth Mindset. Results for the entire sample (n = 156) are shown below the diagonal. Results for subset of participants with complete data (n = 65) are shown above the diagonal.

*p < .05. **p < .01. ***p < .001.

[2] Although we included items relating to perceptions of teaching effectiveness in the high school survey, these data were not present in the data files. Either these items were not administered to students, or the data were not shared for some unknown reason.

Based on theoretical considerations, we expected students' scores on P-TOI to be negatively related to Evaluative Concern and Help-Seeking as Threatening and positively to Help-Seeking Behavior (MSLQ) and Self-Efficacy. Students were asked to think about the effective instructor (Liked as Instructor) while answering items related to these external variables. Thus, we only use estimated scores from Liked as Instructor item bundle in the regression models.

We regressed the four outcomes (in separate models) onto *EAP* values for the Like as Instructor item bundle and included students' growth mindset as a covariate; all variables were z-scored. This was done to control for any variance in the outcomes that could be explained by students' personal theories about intelligence. For model comparison, we generated identical models that did not include the variable of interest, P-TOI (that is, the models only included Growth Mindset as a predictor). Results show that scores on P-TOI significantly predicted Evaluative Concern, $\chi^2(1) = 12.18$, $p < .001$, $r_{\text{partial}} = -.34$, and Self-Efficacy, $\chi^2(1) = 4.43$, $p = .03$, $r_{\text{partial}} = .22$. However, P-TOI did not predict Help-Seeking Behavior (MSLQ), $\chi^2(1) = 0.27$, $p = .60$, $r_{\text{partial}} = .10$, or Help-Seeking as Threatening, $\chi^2(1) = 0.47$, $p = .49$, $r_{\text{partial}} = -.10$. It is important to note here that the Help-Seeking Behavior items have high measurement error (poor internal consistency), which could partly explain the non-significant effect.

Given the similarity between the Evaluative Concern and Help-Seeking as Threatening items, the strong effect for the former and the null effect for the latter is unexpected. If we limit the analysis to the adult sample with available data ($n = 51$), the effect is significant and comparable to Evaluative Concern ($\beta = -0.31$, SE = 0.13, $p = .02$). This provides additional evidence that the high schools in our sample may be different from average (better teaching, motivated administrators, etc.) and this could potentially be diluting the effect.

Table 17

Regression Estimates for Models Testing Convergent Validity

Variables	Evaluative Concern		Help-seeking as Threatening		Self-Efficacy		MSLQ	
	β	SE	β	SE	β	SE	β	SE
With ITOI								
P-TOI	-0.30***	0.09	-0.07	0.10	0.21*	0.10	0.05	0.10
ITOI	-0.25**	0.09	0.07	0.10	-0.00	0.10	-0.00	0.10
Without ITOI								
P-TOI	-0.33***	0.08	-0.06	0.10	0.21*	0.10	0.05	0.10

Note. P-TOI = Perceived Implicit Theories of Intelligence; ITOI = Implicit Theories of Intelligence; MSLQ = Motivated Strategies Learning Questionnaire. Intercept terms have been removed.

***$p < 0.001$, **$p < 0.01$; *$p < 0.05$.

Figure 13

Scatterplots of Outcome Variables in Study 2

Note. Blue lines represent OLS regression lines and gray areas represent 95% confidence intervals.

Divergent Validity

One might worry that students' responses to P-TOI items simply reflect students' personal theories about intelligence. In this sample, students' overall scores on the P-TOI instrument are not correlated with students' Growth Mindset, $r(103) = 0.05$, $p = .61$. The correlation is a little higher, but not significantly different from zero, for item bundles referencing instructors students liked: $r_{LI}(96) = 0.15$, $p = .15$; $r_{LP}(97) = 0.11$, $p = .29$. Correlations are close to zero for the ones they disliked: $r_{DI}(77) = -0.04$, $p = .72$; $r_{DP}(80) = -0.01$, $p = .92$). As these correlations are low and non-significant, there is lack of evidence to indicate an overlap between the two constructs.

Consequences

Although we do not anticipate many negative consequences resulting from the instrument, it could potentially be misused for instructor evaluation or accountability. Additionally, if there are large discrepancies between instructors' beliefs and practices and students' perceptions, this could be psychologically troubling for instructors. However, the issue here might relate to the construct and not the instrument per se (M. Wilson, 2005). Further research should investigate whether these concerns are valid in real-world settings.

Evidence for Fairness

Even though the items are administered to students, the instructors are the target, and in terms of fairness, it is crucial that the instrument not be used for accountability or in a way that could negatively affect instructors. The instrument is meant to be a descriptive tool that should be used for improving teaching practice and assessing whether there is a mismatch between instructor views and student perceptions.

DIF

Another strand of fairness relates to whether the items are fairly assessing different subpopulations. Differential Item Functioning analysis (hereafter DIF) is an evaluation of whether an item is functioning differently for different groups of responders, holding their location on the construct constant. That is, for two students from different groups at the same estimated location on the construct's continuum, are students from one group more (or less) likely to respond positively to an item? This allows one to test the assumption of measurement invariance and is an important step in assessing the fairness of instruments. The analysis is conducted by dividing the sample into two groups and using one as the reference group that provides baseline performance (Mapuranga et al., 2008).

Since gender and under-represented minority (URM) status are both theoretically relevant demographic variables in this context, we tested whether the items functioned differently for female (n = 48) vs. male (n = 70) and URM (n = 53) vs. non-URM students (n = 59). To test the veracity of the intersubjectivity argument made previously, we also test for DIF between the high school (n = 65) and adult (n = 91) sample, as well as between American (n = 110) and non-American (n = 46) sample. Table 18 presents the results of these analyses. The estimated difference (in logits) is presented for only one subgroup, as the difference is symmetric around zero. That is, if the DIF value for males is -0.1, the value for females will be 0.1, and the absolute difference between the two groups is 0.2 logits. Negative values imply that an item was 'easier' for the subgroup (i.e., participants in the subgroup rated the instructor more positively), and positive values indicate the opposite. The DIF value is akin to an effect size that informs us of the amount of DIF exhibited by an item and whether it is of practical significance.

As with the Mean Square values, there is no standard way to evaluate how large the DIF value must be to be considered troubling. A rule of thumb that is used by the Educational Testing Service (ETS) is to use 0.426 as a benchmark (Paek & Wilson, 2011). That is, if the DIF is equal to or larger than an absolute value of 0.426, we can claim the presence of moderate DIF. Since this value is in log-odds (logit), we can exponentiate and interpret it as odds (1.53 or 1.53:1). This means that holding the level of construct constant, members of one group have 53% higher odds of endorsing the item. However, as with everything, the context will determine whether this threshold should be shifted. Given that ours is an attitudinal instrument with stakes lower than in standardized achievement testing, the ETS benchmark is likely conservative, and we recommend moving the threshold higher to 0.69 (2:1 odds of endorsing an item) to be considered worrisome.

DIF analysis reveals one item, Growth Opportunities, with substantial differences between the subgroups (based on the 0.69 criterion), ranging from 0.70 to 1.22 logits (2:1 to 9:1 odds of endorsing an item). This item asks whether the instructors provide opportunities to show growth (e.g., resubmit assignments). The item shows higher DIF for Dislike as Instructor item bundle and for high school and URM subgroups. High school students are more likely to endorse the item (since negative DIF values indicate that the item is 'easier' to agree with), whereas non-U.S. students are more likely to disagree with the item. Thus, this item shows differential functioning for students in high school and students outside the United States. Given that the items inquired about an important aspect of classroom culture, and since the following administration of the instrument was at an American university, we retained the item in the next iteration.

There are 18 other instances of DIF larger than 0.426, but no items except Growth Opportunities show moderate DIF more than thrice. Liked as Instructor item bundle shows the highest amount of DIF (9), and Like as Person, the lowest (3). Female vs. male analysis shows the least difference (3), and non-U.S. vs. U.S. shows the highest (9). Along with the DIF analysis, we also conducted t-tests to assess whether mean scores for the four item bundles differed significantly between subgroups included in the DIF analysis (results in Table 19.1 and 19.2). URM students did not differ significantly from non-URM students on any of the item bundles. Females were more likely than males to rate instructors they Disliked as Instructors lower on P-TOI. High-school students tended to rate their instructors more positively than the adult students and non-U.S. participants rated their instructors more poorly than the U.S. sample. Overall, it appears that the instrument is functioning slightly differently for high-school and non-U.S. participants, although that could be a function of the sample used in this study. Given that only one item shows DIF according to our criteria, and there is minimal DIF for female and URM participants, we do not recommend any changes to the instrument at the moment and recommend further research on larger, younger, and global samples.

Table 18

Differential Item Functioning Results in Study 2

Item	Liked as Instructor				Disliked as Instructor				Liked as Person				Disliked as Person			
	Female	URM	High-School	Non-U.S.	Female	URM	High-School	Non-U.S.	Female	URM	High-School	Non-U.S.	Female	URM	High-School	Non-U.S.
Do Well	-0.09	0.09	0.06	-0.02	-0.02	0.10	0.13*	-0.12	0.05	0.02	-0.09	0.04	-0.14*	0.04	-0.06	-0.02
Improve Abilities	-0.05	-0.12	0.12	0.03	-0.04	0.08	_0.23*_	_-0.24*_	_-0.26*_	-0.09	-0.01	-0.01	0.07	0.08	0.07	-0.08
Growth Possible (Quantitative)	_-0.33*_	-0.15	-0.21*	_0.30*_	0.04	0.09	0.06	-0.04	0.21*	0.04	0.04	0.01	-0.01	0.00	-0.06	-0.04
Growth Possible (Qualitative)	-0.15	-0.02	-0.06	0.08	0.03	0.15*	0.21*	-0.18*	0.02	0.01	0.09	-0.02	0.20*	_0.30*_	0.21*	_-0.26*_
Teaching Care	0.08	0.12	-0.07	-0.12	-0.03	0.14*	-0.03	0.02	-0.10	0.01	-0.03	0.09	-0.15*	-0.11	-0.18*	0.08
Actionable Feedback	-0.10	-0.15	0.05	0.08	-0.08	-0.15*	-0.19*	-0.17*	-0.01	-0.03	-0.06	-0.10	-0.01	-0.03	0.09	-0.09
Growth Opportunities	_0.25*_	_-0.31*_	__-0.59*__	0.02	0.16	__-0.35*__	__-0.99*__	__1.11*__	__-0.40*__	0.18	-0.06	__0.41*__	-0.06	__-0.43*__	__-0.48*__	__0.41*__
Comfort Questions	0.12	-0.07	0.11	-0.09	0.06	-0.01	_0.29*_	-0.21*	0.19*	0.10	0.16*	-0.21*	0.14	0.12	_0.26*_	-0.10
Comfort Comments	0.18*	-0.10	0.13	-0.10	-0.03	0.04	0.15*	_-0.23*_	0.17*	-0.03	0.09	-0.12	0.17*	0.07	0.13	-0.05
Encourage	0.10	_0.25*_	0.04	_0.24*_	-0.04	0.00	-0.09	-0.04	0.10	-0.02	-0.04	-0.05	-0.04	-0.02	-0.10	0.13
Mistake Response	-0.07	_0.24*_	-0.04	-0.05	-0.01	-0.06	-0.02	0.08	0.00	0.12	0.11	-0.11	-0.14*	-0.11	-0.07	0.15*
Normalize Failure	-0.09	0.06	0.20*	_-0.32*_	-0.03	-0.01	_0.27*_	-0.11	-0.02	-0.21*	-0.21*	0.07	0.07	0.13	_0.23*_	-0.21*
Teaching Improve	0.16	0.17	0.26	-0.06	-0.01	-0.02	-0.02	0.12	0.05	-0.11	-0.02	-0.01	-0.12	-0.02	-0.05	0.07

Note. Items that show statistically significant DIF are denoted with "*"; items that show DIF larger than 0.69 are in bold, and larger than 0.42 are italicized. Since the difference (in logits) between the subgroups are symmetric around zero, the criterion value is twice the value shown in the table (i.e., .10 in the table represents an absolute DIF of .20 between the two subgroups).

Table 19.1

Means, Standard Deviations, and T-Test Results for Subgroups

	Female		Male		*t (df)*	*p*	URM		Non-URM		*t (df)*	*p*
	M	*SD*	*M*	SD			*M*	*SD*	*M*	*SD*		
Liked as Instructor	4.23	0.49	4.18	0.54	-0.41 (99)	.68	4.24	0.49	4.18	0.56	-0.58 (97)	.56
Disliked as Instructor	2.97	0.87	3.44	0.92	2.40 (71)	.02	3.09	0.87	3.31	0.97	-1.04 (73)	.30
Liked as Person	4.27	0.42	4.20	0.63	-0.67 (101)	.51	4.27	0.75	4.07	0.54	1.55 (86)	.13
Disliked as Person	2.98	0.91	3.11	1.13	0.61 (81)	.55	3.09	1.03	2.87	1.07	-0.94 (79)	.35

Note. URM = Underrepresented Minority Status

Table 19.2

Means, Standard Deviations, and T-Test Results for Subgroups

	HS		Adult		*t (df)*	*p*	U.S.		Non-U.S.		*t (df)*	*p*
	M	*SD*	*M*	SD			*M*	*SD*	*M*	*SD*		
Liked as Instructor	4.34	0.42	4.11	0.57	-2.46 (108)	0.02	4.32	0.48	4.01	0.53	3.00 (62)	.003
Dislike as Instructor	3.56	0.88	2.78	0.73	-4.72 (92)	<.001	3.44	0.85	2.64	0.78	4.33 (49)	<.001
Liked as Person	4.23	0.68	4.18	0.57	-0.49 (110)	.62	4.27	0.61	4.07	0.64	1.70 (77)	0.09
Disliked as Person	3.49	1.06	2.53	0.75	-5.02 (81)	<.001	3.32	1.02	2.38	0.76	4.89 (72)	<.001

Note. HS = High School; U.S. = United States

Instrument Modifications

Feedback after the initial conceptualization indicated that the original construct map included perceptions of instructor mindset as well as seeking/being receptive to feedback, and both need not co-occur. Thus, the construct map was amended to include only student perceptions. The content of the map was also updated to reflect the insights from students' open-ended responses. Before using the instrument in the next study, we removed five items to truncate the length of the instrument to account for limited survey administration time. The items were removed based on several considerations, including item misfit and potential confounding.

The Encourage item was highly correlated with other items (had the lowest MNSQ values) and was thus removed. Although the psychometric properties of items related to feeling comfortable asking questions or making comments in class were not concerning, there was a higher likelihood of them being confounded with personality characteristics (like extraversion), cultural factors, or other features of the classroom environment not related directly to the instructor. Both items were removed. Items Growth Possible (Quantitative) and Growth Possible (Qualitative) were quite similar in terms of content and correlated at an average of $r = .64$. Thus, to make the instrument less repetitive, we retained Growth Possible (Quantitative), which had better psychometric properties (MNSQ values closer to 1) and item responses that better adhered to Guttman scaling. We removed Teaching Improve in the next administration since the scale was administered to college students at a research university where direct feedback from students during a course is less likely.

Upon reducing the number of items from thirteen to eight, we used the Spearman-Brown prophecy formula (Brown, 1910; Spearman, 1910) to predict changes in instrument reliability. The reliability (Cronbach's α) for Liked as Instructor (lowest reliability in this sample) is expected to decrease from .90 to .85, and the reliability for Disliked as Instructor (highest reliability) is expected to decrease from .96 to .94. Thus, the instrument expected to retain sufficient internal consistency.

Figure 14

Perceived Theories of Intelligence Construct Map (V2)

Respondent	Responses
High Growth: Respondent perceives the instructor to hold the mindset that all students are capable of significant growth; perceives instructor to be invested in every student's learning; feels challenged but supported; is not afraid of being judged for making mistakes or struggling in course.	**High Growth:** "Instructor challenges student thinking, expects everyone to do well, and offers helpful strategies." "I feel really comfortable asking questions in class or asking for help." "Instructor turns student mistakes into learning experiences and normalizes failure." "Instructor is sure people will learn and improve in the class."
Growth: Respondent perceives the instructor to hold the mindset that many but not all students are capable of growth; perceives instructor to be invested in some students' intellectual growth.	**Growth:** "Instructor thinks that some students won't succeed in class, but most can if they put in effort." "Instructor thinks people will improve in class." "I feel comfortable asking for help."
Mixed: Respondent perceives the instructor to hold the mindset that only some students are capable of growth.	**Mixed:** "Instructor does not really care if students improve or not." "I don't really like asking questions in class." "I am usually unsure about asking for help."
Fixed: Respondent perceives the instructor to hold the mindset that only a few students are capable of marginal growth; perceives the instructor to strongly favor those with higher baseline ability; worries about making mistakes.	**Fixed:** "Instructor thinks not all students will be able to excel in the class." "I don't feel comfortable doing to office hours or ask questions."
High Fixed: Respondent perceives the instructor to hold the mindset that students are either smart or they're not; feels incompetent, intimidated, and worries about being judged for making mistakes or getting things wrong.	**High Fixed:** "Instructor encourages students to drop the course if they're struggling." "Instructor only focuses on the "smart" students." "I'm scared to ask for help." "I will look stupid if I ask questions in class."

Discussion

The aim of this study was to create and assess the psychometrics properties of an instrument for measuring students' perceptions of their instructors' mindsets about intelligence. We hypothesized Perceived Implicit Theories of Intelligence (P-TOI) as a unidimensional construct with five levels, and that hypothesis was supported by empirical evidence. Under the assumption that the validity evidence presented in this report is convincing, the results suggest that students are more likely to rate instructors whom they like or consider effective teachers as having an incremental view of intelligence and those they disliked or consider less effective as holding a more entity or fixed view of intelligence.

The item fit statistics indicate that the data fit relatively well to the partial credit model. The Like as Instructor shows the poorest fit to the data, which is likely a function of the way the items were bundled. Since the instrument is not going to be used primarily with effective or ineffective teachers, this misfit is not indicative of how the instrument would work in a real setting. Based on the Wright maps, it appears that the items functioned better for instructors whom the students consider to be less effective or less likable as instructors. This is likely due to higher variance in the responses to these items, also indicated by the observation that the data are spanning more of the hypothesized construct and no levels are missing. The Disliked as Instructor data have the highest spread. As M. Wilson points out, echoing Tolstoy (2002), "Good instructors are all alike; bad instructors are bad in their own way" (personal communication, July 26, 2023). Although it is possible that these instructors are giving out stronger cues that reflect their opinions on the malleability of students' abilities, we don't have enough evidence to support that conjecture.

The correlations between students' own growth mindset and their perceptions of their instructors' mindset do not indicate a spillover of one's own's beliefs about intelligence and hence provide evidence for divergent validity. The correlations are stronger and significant for teachers they liked/thought were good teachers. However, correlations are not high enough to warrant concern that the instrument is simply measuring students' own theories about intelligence. The Like as Instructor and Like as Person item bundles were significantly correlated with Faculty Growth Mindset (Muenks et al., 2020) items, providing evidence for convergent validity.

Evidence from analysis of other external variables shows that the scores on the instrument correlate, as expected, negatively with Evaluative Concern and positively with Self-Efficacy. However, the results are non-significant for variables related to help-seeking. This could be partly due to the higher measurement error in those variables. The relationship between P-TOI and help-seeking is one of the aims of this line of research and not an established finding, so we are less able to draw conclusions about the validity of the instrument from this result. We observe DIF for high school students, and although it could be a function of the sample in our study (schools with better teachers self-selected into the study), it could also be the case the items or the construct function differently for younger students. There is also some evidence of DIF when comparing American and non-American students; further study with larger samples is warranted before drawing any conclusions. Currently, our argument for intersubjectivity is not supported as many items show moderate to large amount of DIF for different samples.

As we mentioned earlier, K-12 teachers might be able to gauge their students' abilities accurately, but similarly, students might also be able to pick up cues from the teachers in these contexts due to closer, protracted interaction. Thus, P-TOI might be better suited for younger students, and since we see DIF between high school and college students, it is possible that P-TOI manifests differently for younger students. These predictions should be studied further, and we recommend some other future directions, as well as limitations, in the next section.

Limitations and Future Directions

A major limitation of this study is the small sample of participants as well as items. This is especially relevant since the instrument did not include items that were very ‘difficult’ to agree with, and we did not have items that covered the higher end of the person distribution (except for Disliked as Instructor). Future studies should include more items that are harder to endorse and recruit a large and diverse sample. A higher-powered study would also allow researchers to account for country-level differences in the models. Another limitation is that the adult sample did not comprise entirely of students, and some respondents reflected on courses they had taken in the past, which is not ideal and induces a source of noise. In the next chapter, we discuss results from a study where the instrument was administered to students in the context of a particular course, which helps address this limitation.

Another source of limitation is that the data were collected during the COVID-19 pandemic, and around half of the high school students were in virtual classrooms (29 out of 65), which meant less direct contact with instructors and a non-typical learning environment. Further studies should evaluate students in actual classrooms as there are likely stronger cues in the present about how the instructors view intelligence and student ability. It is also likely that the patterns for psychological variables we observe were affected by the atypical global context.

Finally, it would be useful to measure instructors’ mindsets for comparison with students’ perceptions. As the instrument includes instructor behaviors, instructors should self-report their beliefs and teaching practices to assess whether student perceptions align with classroom practices. Testing the magnitude of the mismatch between student perceptions and teacher beliefs would be theoretically and practically interesting, as will an evaluation of whether the mismatch is larger for some student subpopulations.

Conclusion

Overall, the P-TOI instruments’ psychometric properties are promising. Further testing and adjustments should be made based on the recommendations above. Finally, a note of caution. The construct (and the instrument) is not meant to be a value-judgement. That is, we are not claiming that having a malleable view of intelligence is always positive. If an instructor holds a view that intelligence is infinitely malleable (‘everyone can be Einstein’), that is (a) not a view that aligns with neuro-scientific evidence, and (b) is surely to be detrimental to students. Therefore, when introducing the instrument to educators, it must be framed carefully, and the primary purpose of the instrument—improvement to teaching practice and, consequently, student learning—should be made salient and kept in the forefront.

CHAPTER IV

SOCIOPSYCHOLOGICAL EXPERIENCES IN GATEWAY STEM COURSES

Structured Abstract

Background: Researchers have recently begun to assess how instructors' theories about intelligence, and students' perceptions of said theories, impact psychological and academic outcomes in the classroom. How these perceptions form and influence student behavior conducive to academic success remains to be addressed.

Purpose: To assess whether students' perceptions of their instructors' theories about intelligence (Perceived Implicit Theories of Intelligence; P-TOI) correlate with their help-seeking behavior (asking questions, visiting office hours, etc.), level of engagement, worries about being evaluated negatively, and considerations about dropping a STEM course.

Participants: College students (n = 316) enrolled in STEM courses at a large public North American university

Research Design: Repeated measure (one semester) observational study

Data Collection and Analysis: Data were collected between August 2021–April 2022 via Qualtrics (https://www.qualtrics.com) and analyzed using R (R Core Team, 2023), ACER Conquest (Adams et al., 2012), and BEAR Assessment System Software (BASS; Wilson & Sloane, 2000). Hypotheses and analysis plan were preregistered.

Findings: P-TOI did not predict help-seeking behavior; however, students who score higher on P-TOI were less likely to worry about being evaluated negatively, less likely to drop or consider dropping the course, and reported higher course engagement during the initial weeks of the semester. Importantly, these perceptions do not predict help-seeking behavior during the initial weeks of the semester, and we find no evidence of differential impact on female, underrepresented minority, and first-generation college students.

Conclusion: Students' perceptions of their STEM instructors' theories about intelligence predict higher course engagement and lower concerns about negative evaluation and odds of dropping the course or considering dropping a rigorous STEM course.

Sociopsychological Experiences in Gateway STEM Courses

Science is the backbone of a well-informed, technologically advanced society. In recent decades, scientific literacy and preparing students for STEM (Science, Technology, Engineering, and Mathematics) disciplines have become a global priority (Chen, 2013; Kennedy & Odell, 2014). This honorable goal is only feasible if students are actively engaged and interested in STEM. Although improvements in science education and teaching methods appropriate for imparting scientific thinking are paramount (National Research Council, 1996), it is also crucial to address psychological barriers that deter students who may otherwise be inclined to pursue STEM from doing so (Brainard & Carlin, 1998; Foley et al., 2017).

As mentioned in the preceding chapters, beliefs about intelligence and ability play an important role in student motivation. And given that intelligence is considered an important component of scientific aptitude, beliefs about intelligence may be especially relevant for the pursuit of science (Leslie et al., 2015). In disciplines such as physics, which sits atop the hierarchy of scientific disciplines, “ability is considered to be innate” (Penner, 2015, p. 235). This view is especially prevalent in some areas of study like mathematics (which forms the bedrock of scientific study), as evidenced by the common use of the expression (in the U.S.), “I’m not a math person” (Gunderson et al., 2017; Meyer et al., 2015). It has been noted that people are more susceptible to self-fulfilling prophecies in new situations and domains for which they have ill-formed self-perceptions (Jussim, 1990). Coupled with inadequate pedagogy, psychological barriers that undermine students’ willingness to exert the effort required to excel in rigorous scientific disciplines are especially worrisome.

The salience of cognitive ability-related beliefs in STEM disciplines makes students’ beliefs about intelligence especially important for understanding their approach or avoidance of STEM fields. It has been posited that students’ implicit theories about intelligence “play a key role in their math and science achievement” (Dweck, 2008). Although considered extremely important across the globe, STEM education is perceived as being ineffective overall (Pew Research Center, 2015). Given that STEM disciplines will dominate job creation in the modern world, it is imperative to address barriers to advancement in STEM fields. Introductory STEM courses—also known as “gateway” (Kroeper et al., 2022) or “weeder” courses (Be et al., n.d.)—have been hypothesized as sources of bottleneck in the STEM pipeline. These courses are often considered to be specifically designed to “weed out” students lacking the caliber and drive for rigorous areas of study (Chen, 2013; Rattan et al., 2018), and negative experiences in such courses can lead to attrition (C. Good et al., 2012; Mervis, 2010).

The Programme for International Student Assessment (PISA) included an item measuring implicit theories of intelligence in its most recent iteration, and results showed that although the link between an incremental view of intelligence and performance was positive for American students, that association was negative for Chinese students (i.e., those with a fixed view of intelligence were more likely to perform better on the assessment; Gouëdard, 2021). Sun and colleagues (2021) have speculated that the difference between American and Chinese students’ conceptualization of intelligence translates to different opinions about the malleability of intelligence. Although they did not look at cultural differences, Limeri et al. (2020a, 2020b) have similarly pointed out that students’ responses on the Implicit Theories of Intelligence scale (Dweck, 2000) differ as a function of their definition of intelligence. Thus, differences in students’ beliefs about the malleability of intelligence may reflect underlying discrepancies in lay beliefs about the nature of intelligence.

Together, these studies indicate that theories about intelligence, as well as the perceived relationship between intelligence and academic success, are potentially subject to cross-cultural differences. Given that these lay beliefs are transmitted through cultural influences and given that cultures may differ on how important intelligence—as opposed to non-cognitive factors like effort, perseverance, and help-seeking—is considered for scholastic success (Hess et al., 1987), students from some cultures may be more susceptible to the effects of maladaptive mindsets about intelligence on academic performance. The social nature of the constructs of interest in this study (perceived theories of intelligence and help-seeking), thus, makes it especially relevant for a cross-cultural examination. We hoped to assess whether (a) students’ perceptions about the malleability of intelligence differ as a function of exposure to Western culture (non-Asian vs. Asian-American/Asian) and (b) whether perceptions of instructors’ theories of intelligence vary between Asian and non-Asian students. Contingent upon a difference between these groups in how they perceive their instructors' theories of intelligence, we were also interested in testing whether Asian vs. non-Asian background interacts with students' theories about intelligence in predicting academic outcomes like engagement, help-seeking, attrition, and grades.

Current Study

Attrition in STEM fields has been linked to attitudinal factors like motivation, confidence, and beliefs about one's capacity to excel in STEM (Ackerman et al., 2013). The purpose of the current research is to extend the recent findings on meta-lay theories about intelligence and students' psychological experiences in STEM courses. We do this by identifying causal mediators between students' perceptions of STEM instructors' theories about intelligence, on the one hand, and students' performance and persistence in introductory STEM courses, on the other.

Following the research delineated in previous chapters, the present study was designed to test whether students' perceptions of their instructor's theories about intelligence predicted their engagement, persistence, and academic help-seeking in the context of specific classrooms. We assessed students' psychological experience during the initial weeks of a "weeder" STEM course at a large public research university. Below, we present our predictions as five research hypotheses, which we tested in this empirical investigation, along with several exploratory analyses.

H1a: Students who perceive the instructor of a gateway STEM course to hold a malleable view of intelligence (those who score higher on the Perceived Theories of Intelligence instrument) will be less likely to drop or consider dropping the course.

H1b: Students who perceive the instructor of a gateway STEM course to hold a malleable view of intelligence will score lower on Evaluative Concern.

H1c: Students who perceive the instructor of a gateway STEM course to hold a malleable view of intelligence will score higher on Course Engagement.

H1d: Students at risk for attrition based on demographic variables (underrepresented minority, female, and first-generation college students)—hereafter, 'at-risk students'—will experience lower course engagement when they perceive the STEM instructor to hold a fixed view of intelligence. (Said differently, the hypothesized effect from H1c will be stronger for students at greater risk of attrition based on demographic variables.)

H2: Students who perceive the STEM instructor to hold a malleable view of intelligence will engage in more academic help-seeking (asking questions in class, going to office hours, emailing the instructors), controlling for the level of academic difficulty faced by the student.

Method

Participants

We targeted students enrolled in rigorous, introductory STEM courses at a large public research university on the west coast of the United States. Interpretations of the findings should consider that our sample was a non-random sample of convenience. Despite this major limitation, the sample potentially allows for insights from a unique, intellectually challenging context purported to "weed out" students from STEM disciplines (Chen, 2013). Given the lay notion that one needs to be “brilliant” to do well in STEM (Bian et al., 2018), we reasoned that theories about intelligence would be especially salient in these contexts. The courses were selected using the following criteria (Be et al., n.d.):

1. Large class size
2. Large number of enrolled students interested in pursuing a major that specifically requires successful completion of the course
3. High rigor to determine whether one can successfully continue in the major
4. High attrition from the course, and consequently out of specific STEM majors

Based on these three criteria, we targeted nine courses from Biology (2), Chemistry (3), Math (2), and Physics (2). We recruited eligible participants via (a) emails to course instructors with requests to share the study flier and website with students (one instructor also allowed the first author to make an announcement in class), (b) the research participant pool of the university’s psychology department, and (c) direct recruitment through flyers around campus and online forums for university students. Participation was limited to students currently enrolled or waitlisted in one of the introductory STEM courses. Students enrolled in more than one target course were asked to answer the surveys in reference to the course they expected to be most challenging.

We collected as many participants as possible across two semesters (Fall 2021–Spring 2022). Students who signed up through the research participation pool received course credit, and everyone else received prorated compensation (up to \$8) for the number of surveys they completed (\$2 each for the first and last surveys and \$1 for each of the four weekly surveys). Additionally, participants who completed all six surveys received a bonus of \$2 for a total compensation of \$10. Participation across the entire study (one semester) was not expected to exceed 60 minutes.

The sample includes 253 females (79.1%) and 53 males (16.6%), and the mean age is 19.2 years ($SD = 1.26$). For other demographic information, please refer to Table 20. Since no research has yet linked the constructs of interest in this study—perceived implicit theories of intelligence and academic help-seeking—it was difficult to estimate an expected effect size a priori. Based on effect sizes typical in this research domain (i.e., implicit theories of intelligence), we preregistered the smallest effect size of interest as $r = .10$ (SESOI; Anvari & Lakens, 2021). However, post-hoc power analysis (using ‘*pwr*’ package in R; Champely, 2020) for a χ^2 test with 1 degree of freedom (our preregistered test for all confirmatory hypotheses) and $\alpha = .05$ revealed that we were 80% powered to find an effect size of at least $r = .16$ for the primary analyses, and only 48% powered to find our smallest effect size of interest ($r = .10$). An important caveat, however, is that due to covariate missingness, not all hypotheses were tested on the entire sample.

Table 20

Sociodemographic Characteristics of Participants in Study 3

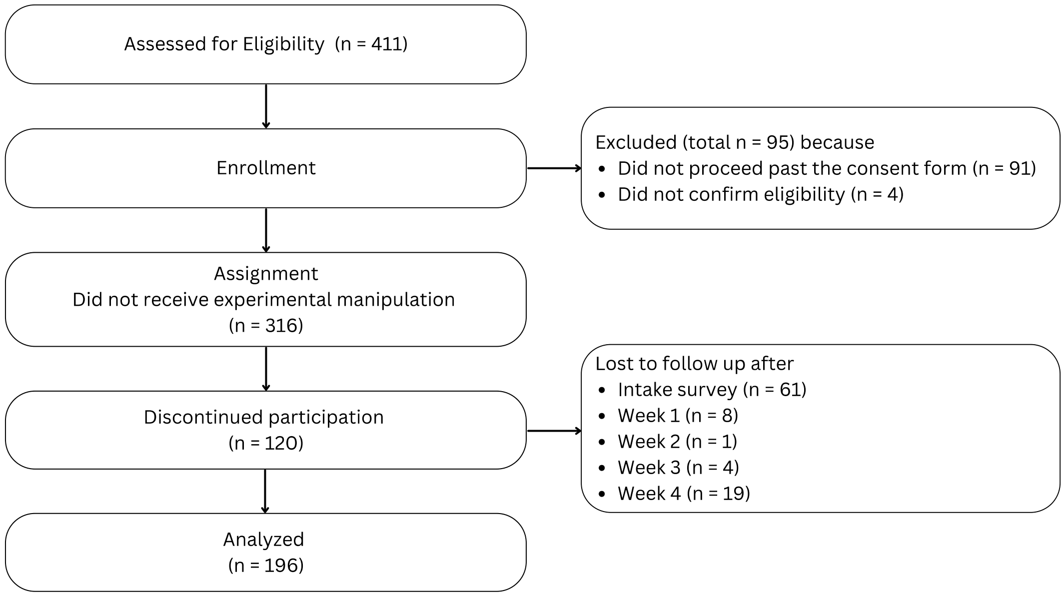
	Biology (N = 59)	Chemistry (N = 109)	Mathematics (N = 72)	Physics (N = 76)	Overall (N = 316)
Age	19.3 (0.86)	19.0 (1.14)	18.8 (1.56)	19.7 (0.96)	19.2 (1.20)
Gender					
Female	48 (81.4%)	90 (82.6%)	53 (73.6%)	58 (76.3%)	249 (78.8%)
Male	7 (11.9%)	18 (16.5%)	16 (22.2%)	12 (15.8%)	53 (16.8%)
Non-binary/Other	1 (1.7%)	1 (0.9%)	0 (0.0%)	3 (3.9%)	5 (1.6%)
Missing	3 (5.1%)	0 (0.9%)	3 (4.2%)	3 (3.9%)	9 (2.8%)
Race/Ethnicity					
Asian/Asian-American	32 (54.2%)	66 (60.6%)	31 (43.1%)	39 (51.3%)	168 (53.2%)
Black/African-American	1 (1.7%)	1 (0.9%)	1 (1.4%)	0 (0%)	3 (0.9%)
Hispanic/Latinx	8 (13.6%)	13 (11.9%)	13 (18.1%)	10 (13.2%)	44 (13.9%)
White/European	11 (18.6%)	14 (12.8%)	14 (19.4%)	14 (18.4%)	53 (16.8%)
Biracial/Mixed	2 (3.4%)	7 (6.4%)	7 (9.7%)	7 (9.2%)	23 (7.3%)
Other	2 (3.4%)	7 (6.4%)	2 (2.8%)	2 (2.6%)	13 (4.1%)
Missing	3 (5.1%)	1 (0.9%)	4 (5.6%)	4 (5.3%)	12 (3.8%)
Year in College					
Freshman	12 (20.3%)	46 (42.2%)	55 (76.4%)	9 (11.8%)	122 (38.6%)
Sophomore	36 (61.0%)	50 (45.9%)	5 (6.9%)	23 (30.3%)	114 (36.1%)
Junior	8 (13.6%)	7 (6.4%)	6 (8.3%)	31 (40.8%)	52 (16.5%)
Senior	0 (0.0%)	6 (5.5%)	2 (2.8%)	10 (13.2%)	18 (5.7%)
Missing	3 (5.1%)	0 (0.0%)	4 (5.6%)	3 (3.9%)	10 (3.2%)

Inclusion and Exclusion

To be eligible for the study, students had to be enrolled or waitlisted in one of the target STEM courses at the time of study intake; this criterion led to the exclusion of four students. All participants were required to confirm that they were at least 18 years old. We did not preregister any other inclusion criteria. During analysis, we added an additional criterion and excluded participants who did not proceed beyond the consent form and eligibility screening (n = 91). Our final sample includes 316 students, although we have data for the primary independent variable for only a subset (n = 196). We retained the full sample for descriptive and exploratory analyses. For the (exploratory) longitudinal analyses, our preregistered plan was to include participants who completed at least two weekly surveys in addition to the intake and exit surveys (4–6 data points), which resulted in a sample size of 155. Some target courses were taught by multiple instructors; because our instruments asked participants to report on a single instructor, we present supplementary analyses that excluded participants enrolled in these courses (excluded n = 69) in Appendix D. Importantly, the pattern of results for this subset did not differ from that obtained from the full sample.

Figure 15

CONSORT Flowchart of Participants in Study 3



Procedure

This was a semester-long observational study consisting of one intake survey, four short weekly surveys, an exit survey, and a follow-up survey. The intake survey, completed before or during the first official week of instruction, included the consent form, baseline questionnaires, and demographic items. The short weekly surveys were administered during the first four full weeks of the semester and included items pertaining to students' help-seeking behavior, self-efficacy, engagement, study habits, concerns about being evaluated negatively, and whether they had dropped the course, thought about dropping the course, or changed their grading options. In the exit survey, administered at the end of the semester, students reported their perceptions about the course and the instructors. The follow-up survey was sent a month later to procure students' official grades in the course.

Instruments

Variables of interest in the confirmatory hypothesis testing are introduced below; variables that pertain only to exploratory analyses are presented in the exploratory section. A list of all administered instruments/items, along with time the of administration, is available in Table 22. (Complete surveys are available in Supplementary Material).

Perceived Implicit Theories of Intelligence

Perceived Implicit Theories of Intelligence (P-TOI) is the primary independent variable and was measured using an 8-item instrument described in the preceding chapter. The same measurement model—the partial credit model (Masters & Wright, 1982)—was used in the current analysis. The instrument was administered twice over the course of the study. At the beginning of the semester, we administered the P-TOI instrument and asked students to report their general impression of STEM instructors at the institution (P-TOI [G]; students in their first semester in college were excused from answering these items). At the end of the semester, students were asked the same items in reference to the STEM instructor of the specific course they were currently enrolled in that semester (P-TOI [S]). We structured the data collection this way since P-TOI items inquire about instructor behaviors that students would not have been able to evaluate at the beginning of the semester. When students' baseline perceptions of STEM instructors' theories about intelligence (P-TOI [G]) is included as a covariate in the confirmatory hypothesis testing models, the results do not change (see Appendix D for details).

The first step in data analysis was confirming the unidimensionality of P-TOI and ensuring that the data reflected the five hypothesized levels of the construct. The data fit the model well, and the *EAP* reliability was .83 for P-TOI (G) and .81 for P-TOI (S). Only two items misfit slightly (see Appendix D for details); however, as none of the weighted mean square (*MNSQ*) values lay outside the preregistered bounds of 0.5–1.6, all items were retained in the calibration that provided the composite scaled scores (person *EAP* estimates). Overall, the score distributions were left-skewed, both for STEM instructors at the university in general (P-TOI [G]; $M = 3.51$, $SD = 0.62$, range = 1.00–4.63) and for the STEM instructor whom they were taking a course with that semester (P-TOI [S]; $M = 3.82$, $SD = 0.60$, range = 1.38–4.63). (Note that the Guttman scale was equivalent to a numerical range of 1–5.)

In this study, students answered the P-TOI (S) items in reference to a specific instructor, thus providing an opportunity for gathering additional validity evidence for the instrument. To that end, we collected students' responses to the Faculty Growth Mindset items (Muenks et al., 2020), which measure the same construct as P-TOI, during the final follow-up survey. Students' responses on Faculty Growth Mindset items were moderately correlated with P-TOI (S), $r(170) = .58$, 95% CI [.47, .67], $p < .001$. We expected the correlation to be smaller with items about students' impressions of faculty in general (i.e., P-TOI [G]), and that appears to be the case, $r(120) = .32$, 95% CI [.15, .47], $p < .001$.

Thus, for these data, the P-TOI instrument exhibits acceptable psychometric properties and is significantly correlated with another instrument measuring perceptions of instructors' theories about intelligence.

Implicit Theories of Intelligence

Students' own perceptions about the nature of intelligence (malleable vs. fixed), i.e., their Implicit Theories of Intelligence (ITOI), were measured using three items during the intake survey. The items were rated on a 6-point Likert scale from 1 = *strongly agree* to 5 = *strongly disagree* ($M = 4.20$, $SD = 1.15$), and the partial credit model (Masters, 1982) was used as the measurement model (*EAP* reliability = .96).

Academic Help-Seeking

Academic help-seeking was measured during the weekly surveys using eight dichotomous (1 = *yes*, 0 = *no*) items that inquired about students' help-seeking behavior during the week. Items assessed whether students sought help from instructors, teaching assistants, peers, or tutoring services on campus (see Supplementary Material for the entire set). A composite was created by summing the item scores for each week and taking the average across four weeks ($M = 7.21$, $SD = 5.47$; range = 0–31).

Course Engagement

Course Engagement refers to self-reported engagement in the course, assessed using a single item on the weekly survey: “This week for this course (including lecture, lab, and section), I felt...” with response options ranging from 1 = *not at all interested/motivated* to 5 = *extremely interested/motivated* (M = 2.70, SD = 0.81). Although this construct was preregistered to be a composite of self-reported engagement, time spent studying outside of class, and lecture/section attendance, the composite showed extremely poor internal consistency (Cronbach’s α = .24) and was thus abandoned in favor of the single item.

During the intake survey, we collected student ratings on a single item to measure baseline engagement: "Typically, I participate actively (ask/answer questions in class and section, make comments, go to office hours, email instructors, etc.) in the courses I take." However, prior to data analysis, it was deemed that the item content overlapped considerably with our operational definition of help-seeking and would be more appropriate as a measure of *baseline help-seeking* instead. This item is used only for exploratory analysis.

Evaluative Concern

Evaluative Concern taps into feelings of anxiety about being judged as unintelligent and is a composite based on a) perception of academic help-seeking as threatening (3 items) and b) evaluative concern (3 items). The composites were created separately for each week (also created using a partial credit model; EAP reliabilities range from .88–.91), and EAP estimates were averaged across four weeks.

Analysis Plan

Hypotheses and analysis plan were preregistered (https://osf.io/ge2zc). Results from all confirmatory hypotheses are reported here. Following recent recommendations for ‘blinded’ data analysis to avoid tilting the scales in favor of significant results (Dutilh et al., 2019; MacCoun & Perlmutter, 2015, 2017), we included a blinding procedure prior to observation of raw data (https://osf.io/46ucj).

Data Blinding

Researchers often search for mistakes only when the results disconfirm their hypotheses, which creates a bias in favor of significant results (MacCoun & Perlmutter, 2015). A blinding procedure allows the analyst to consider any unexpected peculiarities of the data during the analysis while simultaneously avoiding confirmation bias as the true results remain unknown (Dutilh et al., 2019). In the current study, following Dutilh et al.’s (2019) recommendation for multiple regression, the values of key dependent variables for the primary analyses were shuffled (permuted). If the outcome was a composite based on a measurement model, the permutation was conducted on the original items before the composite was created. Following is the list of dependent variables that were blinded: academic help-seeking, course engagement, dropping/withdrawing or grade change, evaluative concern, and course grade.

After the dependent variables were randomly permuted, we conducted the preregistered analyses on the blinded dataset. Once the analysis was coded, the code was submitted to the public OSF repository (https://osf.io/jwuyq/). The unblinded dataset was analyzed using the registered code only after the analysis plan was preregistered. Although the study had already been preregistered prior to the adoption of the blinding procedure, and the primary analyst (the author) was also the data manager, the blinding procedure was followed faithfully, with the intention of minimizing confirmation bias during the analysis.

Table 21

Instruments/Items Included in Study 3

Instrument	Cronbach's α	Description	Intake	W1	W2	W3	W4	Exit
P-TOI (General)	.83	8 items from the study in Chapter 3; asked in relation to STEM instructors at UC Berkeley in general. Example item: "In general, STEM professors at UC Berkeley think that…" *Response options*: 1 = *…no students are capable of growth* to 5 = *…all students are capable of growth*	X					
P-TOI (Specific)	.82	Same 8 items as P-TOI (General); asked in relation to STEM instructor of the course students were enrolled in that semester.						X
Growth Mindset (ITOI)	.92	3-item instrument that assesses beliefs about the malleability of intelligence *Source*: Dweck (2000); Yeager et al. (2016) *Response options*: 6-point Likert scale from 1 = *strongly agree* to 6 = *strongly disagree* *Example item*: "Your intelligence is something about you that you can't change very much."	X					
Perceived Faculty Growth Mindset	.80	4-item instrument that measures students' perceptions of their instructors' beliefs about the malleability of intelligence *Source*: Muenks et al. (2020) *Response options*: 6-point Likert scale from 1 = *strongly agree* to 6 = *strongly disagree* *Example item*: "In general, most professors at my institution seem to believe that students have a certain amount of intelligence, and they really can't do much to change it."						X
Academic Help-Seeking	.66–.75	8 items *Source*: Ad hoc *Example item*: "This week for this course, did you go to the instructor's office hours to ask questions about the course content?" *Response options*: 1 = *yes, 0 = no*		X	X	X	X	
Help-Seeking as Threatening	.87–.94	3 items that measure students' concerns about being assessed negatively by instructors and peers for seeking help *Source*: Skaalvik & Skaalvik (2005) *Response options*: 5-point Likert scale from 1 = *false* to 5 = *true* *Example item*: "If I ask for help, the professor or [TA] will think I am stupid."	X	X	X	X	X	

Instrument	Cronbach's α	Description	Intake	W1	W2	W3	W4	Exit
Evaluative Concern	.89–.92	3-item that asked students whether they worried about their intelligence or abilities being judged negatively by their instructors and peers. *Source*: Muenks et al. (2020) *Response options*: 7-point Likert scale from 1 = *not at all* to 7 = *very much* *Example item*: "This week for this course... how much did you worry about coming across as unintelligent to your instructor ?"		X	X	X	X	
Baseline Self-Efficacy	.96	8-item subscale from the Motivation Scales (Expectancy Components) section of The Motivated Strategies for Learning Questionnaire (MSLQ) that measure students' confidence in their ability to well in the course (Self-Efficacy for Learning and Performance) *Source*: Pintrich & DeGroot (1990) *Example item*: "I'm confident I can do an excellent job on the assignments and tests in this course." *Response options*: 7-point Likert scale from 1 = *not at all true of me* to 7 = *very true of me*	X					
Course Self-Efficacy	.93–.95	3 out of the 8 items from Baseline Self-Efficacy *Source*: Pintrich & DeGroot (1990) *Example item:* "I'm confident I can understand the most complex material in this course." *Response options*: 7-point Likert scale from 1 = *not at all true of me* to 7 = *very true of me*		X	X	X	X	
Course Engagement		"This week for this course (including lecture, lab, and section), I felt..." (1 = *not at all interested/motivated* to 5 = *extremely interested/motivated)*		X	X	X	X	
Studying		"This week, how many hours did you spend on this course, not including time spent in lecture and section)?"		X	X	X	X	
Baseline Help-Seeking		"Typically, I participate actively (ask/answer questions in class and section, make comments, go to office hours, email instructors, etc.) in the courses I take." (1 = *strongly disagree* to 6 = *strongly agree*)	X					
Academic Difficulty		"This past week, did you struggle academically in this course?" (*Yes/No*)		X	X	X	X	
Effort		"If I work hard, I can do well in this course."	X					
Family Ed-Value		"When it comes to the value of higher education, my family feels..." (1 = *extremely negative about it* to 5 = *extremely positive about it*)	X					

Instrument	Cronbach's α	Description	Intake	W1	W2	W3	W4	Exit
Course-Related								
Enrollment Status		"Are you still enrolled in this course?" (*Yes/No*)		X	X	X	X	
Pass/No-Pass Grading		"Did you change the grading option from letter grade to Pass/No-Pass for this course?" *Follow-up items*: "Why did you choose to change the grading option to Pass/No Pass?"; "How much was this decision influenced by the [instructor's/TA's/classmates'] attitude?"	X	X		X		X
Prerequisites		Have you taken (or are concurrently taking) all the required prerequisites for this course? (*Yes/No)*	X					
Prior Courses		2 items asking students which other courses in the subject students have taken before, where, and what grade they received.	X					
Expected attendance		2 items asking if students plan on attending (or watching) [none/some/around half/most/all] lectures and discussions.	X					
Desired Grade		"What grade would you *like* to get in this course?"	X					
Expected Grade		"What grade do you *expect* to get in this course?"	X	X				
Drop Intention		"Did you consider dropping the course this week?" (*Yes/No*) *Open-ended follow up*: "What made you consider dropping this course?"		X	X	X	X	
Midterm Grade		"Do you have the results from the first midterm (or a major assignment)?" (*Yes/No*) "What grade did you receive on the midterm (or major assignment)?"				X		
Attendance		2 items asking if students attended (or watching) [none/some/around half/most/all] lectures and discussions.						X
Instructor Approachability		2 items: "How approachable, supportive, and helpful did you find your course [instructor / TA]?" (1 = *not at all* to 7 = *very much*)					X	
Course Grade		"What was your official grade in the course?"						X
Demographics								

Instrument	Cronbach's α	Description	Intake	W1	W2	W3	W4	Exit
Age		"What is your age?"	X					
Gender		"What is your gender?" (*Male / Female / Trans or non-binary/Other/Prefer not to say*)	X					
Race/ethnicity		"Which of the following best represents your racial and/or ethnic background?"	X					
Immigration status		"Which country were you born in?' "How old were you when you moved to the United States?"	X					
Major		"What is your intended/declared major?"	X					
Semesters completed		"How many semesters have you *completed* at UC Berkeley? (select 0 if this is your first semester)"	X					
Pre-med status		"Are you pre-med?" (*Yes*/*No*)			X			
First generation status		"Are you the first in your family to attend college?" (*Yes/No*)	X					
Transfer status		Are you a transfer student? (transferred from a two-year institution/community college to UC Berkeley) (*Yes/No*)	X					
GPA		"What is the cumulative GPA from your coursework at UC Berkeley?"	X					
High School GPA		"What is your cumulative high school GPA (non-weighted)?"	X					
Parents' education level		"What is the highest level of formal education completed by your [mother/father]?"	X					
Subjective SES (McArthur's ladder)		MacArthur Scale of Subjective Social Status (Adler et al., 2000); participants were presented with a picture of a ladder with 10 rungs and asked to select where they stand relative to others (in the U.S.)	X					

Note. P-TOI (G) = Perceived Implicit Theories of Intelligence (General); P-TOI (S) = Perceived Implicit Theories of Intelligence (Specific); ITOI = Implicit Theories of Intelligence; GPA = Grade Point Average; SES = Socioeconomic Status.

Primary Analyses

We conducted multiple linear regressions to test hypotheses H1b-d. Although we had preregistered a linear regression to test hypothesis H1a, we deviated from the plan and instead used a multiple logistic regression since the primary outcome variable was binary. Hypotheses were tested using nested model comparison (χ^2 test with $df = 1$); the primary models regressed the dependent variables on P-TOI (S), and the comparison models excluded P-TOI (S). All models controlled for students' implicit theories of intelligence (ITOI), GPA, gender (1 = *female*, 0 = *male/other*), URM (under-represented minority status; 1 = *non-White/non-Asian* , 0 = *White/Asian*), class standing (reference group was freshman), and course subject (reference group was biology).

Open Practice Statement

This research was approved by UC Berkeley's Committee for Protection of Human Subjects, protocol #2021-07-14508. We preregistered the hypotheses, analytic strategy, and tests of confirmatory hypotheses (https://osf.io/ge2zc). The protocol and analysis code for the blind analysis is available at https://osf.io/jwuyq/. Anonymized raw data and code are publicly accessible at https://osf.io/pu8yc/ (Mehta & Bunge, 2021), and survey materials are available at https://researchbox.org/870.

Results

The flow of participants through each stage of the study is depicted in Figure 15. Attrition (operationalized as not responding to the exit survey) did not depend on any baseline characteristics aside from two demographic variables: age and first-generation college student status. For each additional unit of age (in years), students had 22% lower odds of completing the study (OR = 0.78, b = -0.25, SE = 1.02, p = 0.01), and first-generation college students had 46% lower odds of completing the study (OR = 0.54, b = -0.62, SE = 0.26, p = 0.02).

Primary Results

As noted previously, we had five preregistered hypotheses. Below are the results for each of the confirmatory tests (we refer to P-TOI (S) as P-TOI below for convenience):

(a) H1a stated that students who scored higher on Perceived Theories of Intelligence (P-TOI) instrument would be less likely to drop or consider dropping the course during the semester. Corroborating H1a, our results show that, controlling for covariates, each unit increase on P-TOI was associated with 49% lower odds of dropping or considering dropping the course, χ^2 (1) = 9.92, n = 167, p = 0.002 (OR = 0.51, CI [0.32, 0.78], $r_{partial}$ = -.61).

(b) Our second hypothesis, H1b, was that students who score higher on P-TOI would be less concerned about being evaluated negatively by instructors and peers. Consistent with this hypothesis, we found that for each unit increase on P-TOI, students were, on average, 0.25 units less concerned about being evaluated negatively, χ^2 (1) = 7.47, n = 164, p = .005, $r_{partial}$ = -0.22.

(c) Hypothesis H1c stated that students who scored higher on P-TOI would self-report experiencing higher course engagement (H1c). Consistent with this hypothesis, we found that for each unit increase on P-TOI, students reported experiencing, on average, 0.32 units higher course engagement, χ^2 (1) = 15.4, n = 163, p < .001, $r_{partial}$ = 0.33.

(d) We did not find evidence in support of H1d; the interaction between P-TOI (S) and at-risk status was not significant when predicting course engagement, $\chi^2(1) = 0.13$, $n = 163$, $p = 0.69$, $r_{partial} = 0.03$. That is, the relationship between P-TOI and course engagement did not differ between at-risk and not-at-risk students.

(e) Further, scoring higher on P-TOI was not positively correlated with academic help-seeking, $\chi^2(1) = 0.33$, $n = 164$, $p = 0.55$, $r_{partial} = 0.05$. Thus, contrary to our H2 prediction, scores on P-TOI did not predict actual help-seeking behavior during the initial weeks of the course.

In summary, the first three hypotheses were empirically supported: students who scored higher on P-TOI were less concerned about negative evaluation from instructors and classmates, more engaged, and less likely to drop or consider dropping the course. Importantly, these analyses controlled for various student-specific characteristics, including students' own theories about intelligence. We did not, however, find support for hypotheses 1d and 2. The results did not indicate that students at risk for attrition experienced higher course engagement as a function of P-TOI, or that P-TOI was associated with help-seeking behavior during the initial weeks of the semester.

Secondary Results

Along with testing the confirmatory hypotheses using multiple regressions, we preregistered multilevel models to account for the repeated measures design as an exploratory step. As noted in the preregistration, we also planned to explore auxiliary questions for which we did not have explicit predictions. These questions pertain to other factors related to students' experience in the course, namely, their level of subjective motivation, enjoyment, and challenge; how competitive they found the course environment; and how much they felt they had learned from the course compared to other courses. In the next section, we first introduce variables not included in the primary analyses, followed by the results of exploratory analyses.

Secondary Variables

Instructor Approachability

College students, at least in the U.S., might interact more with teaching assistants (TAs) than with faculty, especially in lower-division courses; hence, their experience might be influenced more by interactions with Tas than course instructors. Moreover, a plausible alternative explanation of our results is general liking for the instructor due to perceived helpfulness and approachability, instead of the instructors' perceived theories about intelligence. To help addresses this concern, we collected a single item in the exit survey, administered twice, that asked students how "approachable, supportive, and helpful" (1 = *not at all* to 7 = *very much*) they found their STEM instructor ($M = 5.09$, $SD = 1.51$) and their TA ($M = 5.41$, $SD = 1.57$). TAs were rated as more approachable than course instructors, $t(195) = -2.46$, $p = .01$.

Course Self-Efficacy

Course Self-Efficacy refers to the extent to which students feel confident in their ability to do well in the course. We measured baseline self-efficacy using an 8-item instrument during the intake survey ($M = 4.95$, $SD = 1.29$; Cronbach's $\alpha = 0.96$), and used a partial credit model to get person estimates (*EAP* reliability = 0.99). Three out of the eight items were used during the weekly surveys; we created composites separately for each week (using a partial credit model; *EAP* reliabilities range from .88–.90), and *EAP* estimates were averaged across four weeks.

Table 22

Regression Estimates for Primary Results in Study 3

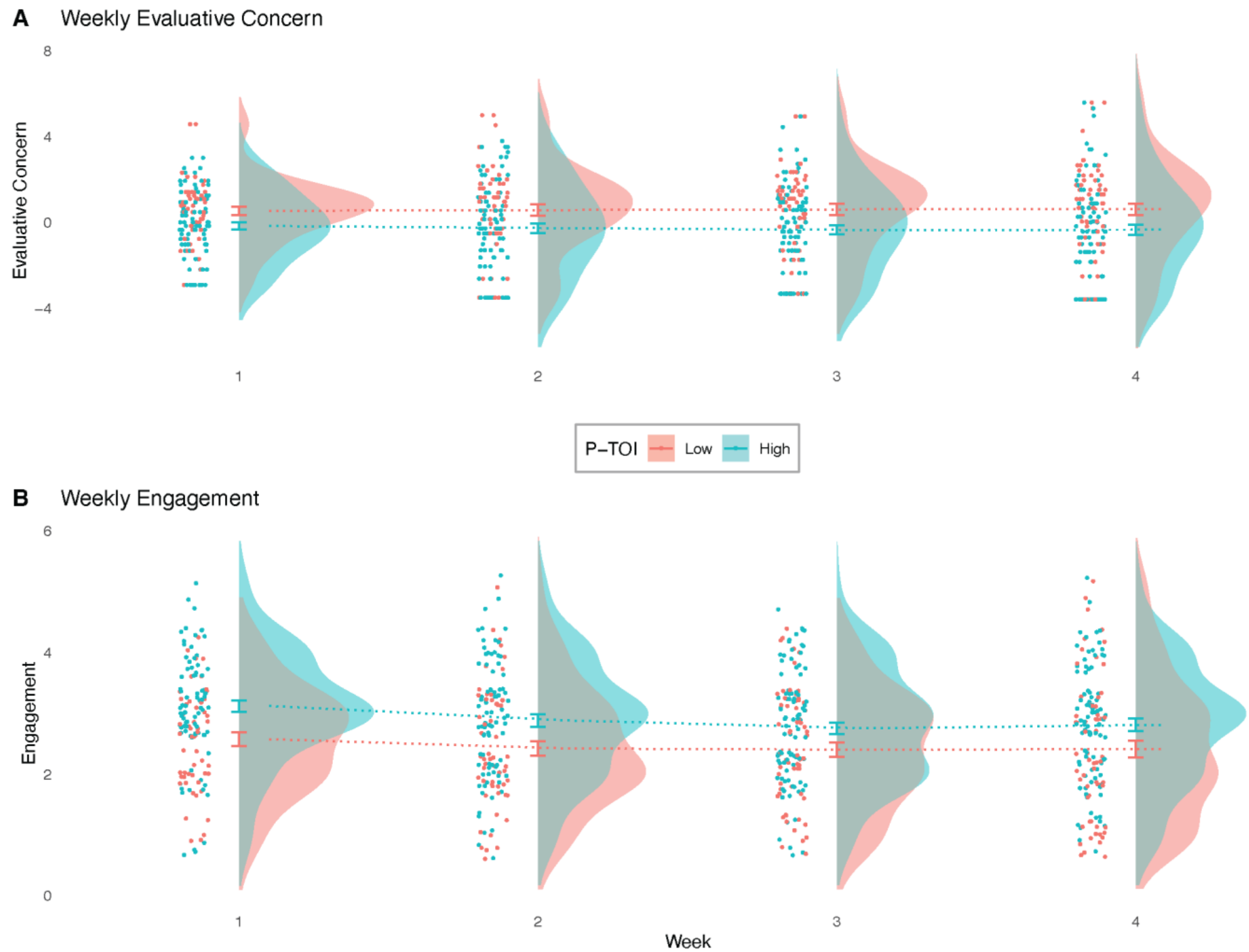
	H1a: Dropping/ Intentions to Drop		H1b: Evaluative Concern		H1c: Course Engagement		H1d: Course Engagement × At Risk		H2: Help-seeking	
	β	SE	β	SE	β	SE	β	SE	β	SE
Intercept	-3.14**	1.12	-0.28	0.28	0.47	0.30	0.49	0.31	0.18	0.29
P-TOI (S)	-0.68**	0.22	-0.22**	0.09	0.32***	0.08	0.40*	0.18	0.05	0.08
ITOI	-0.19	0.23	-0.14	0.09	0.09	0.07	0.08	0.07	0.03	0.08
GPA	-0.03	0.24	-0.15*	0.07	0.04	0.08	0.01	0.08	0.15	0.10
$\text{Subject}_{\text{CHEM}}$	2.23*	0.90	0.16	0.21	-0.28	0.22	-0.27	0.22	-0.00	0.23
$\text{Subject}_{\text{MATH}}$	0.94	0.98	0.08	0.27	-0.08	0.28	-0.08	0.28	-0.24	0.29
$\text{Subject}_{\text{PHYS}}$	1.56	0.99	-0.17	0.21	-0.27	0.23	-0.25	0.23	-0.08	0.26
Female	0.64	0.63	0.10	0.19	-0.31	0.16	—		-0.00	0.16
URM	0.36	0.55	0.26	0.22	0.03	0.19	—		0.02	0.18
$\text{Year}_{\text{SOPH}}$	-0.89	0.53	0.17	0.22	0.12	0.21	0.12	0.21	-0.24	0.20
$\text{Year}_{\text{JUNIOR}}$	-1.35	0.81	0.18	0.29	-0.20	0.27	-0.19	0.27	0.08	0.23
$\text{Year}_{\text{SENIOR}}$	-1.06	1.22	0.37	0.37	-0.46	0.29	-0.43	0.29	0.64	0.58
At Risk	—		—		—		-0.34	0.21	—	
P-TOI × At Risk	—		—		—		-0.09	0.20	—	
Academic Difficulty	—		—		—		—		0.37	

Note. P-TOI = Perceived Theories of Intelligence; ITOI = Implicit Theories of Intelligence; URM = Underrepresented Racial Minority. Female and URM not included in model H1d due to dependence with at-risk status.

***p < 0.001, **p < 0.01; *p < 0.05.

Figure 16

Weekly Evaluative Concern and Engagement in Study 3



Note. P-TOI = Perceived Theories of Intelligence (Specific), factorized (mean split) for visualization. Individual data points are plotted across four weeks; dotted lines represent group means; error bars represent 95% confidence intervals.

Subjective Course Experience

During the exit survey, we asked students to rate their subjective experience in the course on a few dimensions (all measured with single items with Likert response options). Items assessed (1) how challenging students found the course (1 = *extremely easy* to 5 = *extremely challenging*; $M = 3.81$, $SD = 0.79$), (2) how competitive they found their classmates (1= *not at all competitive* to 5 = *extremely competitive*; $M = 2.84$, $SD = 1.23$), (3) how supportive they found their classmates (1= *not at all supportive* to 5 = *extremely supportive*; $M = 3.18$, $SD = 1.23$), and (4) and how much they enjoyed the course (1 = *not at all* to 5 = *immensely*; $M = 2.77$, $SD = 0.93$).

Additionally, we asked students to compare the current STEM course to all their other courses at the university and report (1) whether they felt more or less motivated (1 = *significantly less motivated* to 5 = *significantly more motivated*, $M = 2.77$, $SD = 1.06$), and (2) whether they felt they had learned more or less in the course (1 = *learned significantly less* to 5 = *learned significantly more*, $M = 3.24$, $SD = 0.94$). We also asked how many (1 = *none* to 5 = *all*) lectures ($M = 4.51$, $SD = 0.91$) and TA's discussion sections ($M = 3.84$, $SD = 1.14$) students attended or watched during the semester.

Finally, we asked students to give a subjective rating of how much their experience in the course was influenced (1 = *not at all* to 7 = *very much*) by (1) the instructor's attitude ($M = 5.11$, $SD = 1.64$), (2) the TA's attitude ($M = 5.04$, $SD = 1.69$), and (3) their classmates' attitude ($M = 4.31$, $SD = 1.77$). Students reported that their experience in the course was influenced more by the attitude of the instructor and TA than peers.

Student Experience and Instructor Approachability

Our first exploratory question was whether students' subjective experience in the course depended more on the course instructor(s) or the TA. However, we first looked at attendance as an outcome to assess the validity of the single items about TA and instructor approachability. Consistent with the idea that approachability likely influences attendance, instructor approachability ($\beta = 0.26$, $SE = 0.07$), but not TA approachability ($\beta = 0.08$, $SE = 0.07$) predicted lecture attendance or recording views $F(2, 193) = 9.07$, $p < .001$. Conversely, TA approachability ($\beta = 0.32$, $SE = 0.07$), but not instructor approachability ($\beta = -0.05$, $SE = 0.07$) predicted section attendance or recording views, $F(2, 193) = 10.4$, $p < .001$.

Next, we conducted multiple regression analysis to assess whether students' subjective experience of course challenge, course enjoyment, peer competitiveness, and peer support was predicted by instructor and/or TA approachability. We also evaluated subjective motivation and subjective learning (compared to other courses, how motivated students felt and how much they reported learning in the course) as outcomes in identical models. Standardized regression estimates and standard errors from all six models are presented in Table 24.

Results showed that perceiving the instructor to be more approachable correlated with finding the course less challenging, but TA approachability was unrelated to how challenging students found the course, $F(2, 194) = 10.92$, $p < .001$. Similarly, instructor approachability was associated with lower perceptions of competitiveness, but TA approachability was not related to how competitive students found their classmates, $F(2, 194) = 5.99$, $p = .003$. On the other hand, TA approachability was associated with how supportive students found their peers, but instructor approachability did not predict peer support, $F(2, 194) = 5.67$, $p = .004$. Both instructor approachability and TA approachability positively predicted how much students enjoyed the course, $F(2, 194) = 27.61$, $p < .001$, although the effect was considerably larger for instructor approachability. And lastly, instructor approachability positively predicted both subjective motivation, $F(2, 194) = 12.6$, $p < .001$, and subjective learning, $F(2, 171) = 15.5$, $p < .001$, and TA approachability was not significantly related to either.

Thus, instructor (but not TA) approachability is negatively related to subjective perceptions of challenge and competitiveness in the classroom, and positively related to subjective motivation and learning. TA (but not instructor) approachability is positively related feeling supported by peers. Both instructor and TA approachability are positively related to course enjoyment.

Importantly, we also tested whether P-TOI explained any variance in the subjective student experience outcomes above and beyond approachability. P-TOI (S) and instructor approachability were significantly correlated ($r(194) = .57$, $p < .001$), and when P-TOI (S) was included in the aforementioned multiple regression models, P-TOI significantly predicted all outcomes except peer support. Both P-TOI (S) and instructor approachability, but not TA approachability, predicted how challenging students found the course, $F(3, 193) = 9.65$, $p < .001$. We get similar results for how much students enjoyed the course; both P-TOI (S) and instructor approachability predicted enjoyment of the course, and TA approachability no longer did, $F(3, 193) = 20.2$, $p < .001$. In the model with all three predictors, only P-TOI (S) predicted lower perceptions of competitiveness among peers, $F(3, 193) = 5.41$, $p = .001$. And the experience of peer support was correlated positively only with TA approachability, and not with P-TOI (S) or instructor approachability, $F(3, 193) = 3.77$, $p = .01$.

Table 23

Regression Estimates for Models Predicting Subjective Student Experience

	Challenge		Enjoyment		Competitiveness		Peer Support		Subjective Motivation		Subjective Learning	
	β	SE	*β*	SE	*β*	SE	*β*	SE	*β*	SE	*β*	SE
Not Including P-TOI (S)												
Instructor	-0.28***	0.07	0.41***	0.07	-0.15*	0.07	0.09	0.07	0.34***	0.07	0.33***	0.07
TA	-0.08	0.07	0.14*	0.07	-0.14	0.07	0.19**	0.07	0.00	0.07	0.13	0.07
Including P-TOI (S)												
P-TOI (S)	-0.21*	0.08	0.16*	0.08	-0.17*	0.09	-0.01	0.09	0.26**	0.08	0.21*	0.09
Instructor	-0.17*	0.08	0.32***	0.08	-0.07	0.09	0.10	0.09	0.20*	0.08	0.22**	0.08
TA	-0.04	0.07	0.11	0.07	-0.11	0.07	0.20**	0.07	-0.05	0.07	0.08	0.07

Note. P-TOI = Perceived Theories of Intelligence; ITOI = Implicit Theories of Intelligence; TA = Teaching Assistant. Intercepts have been removed.

*$p < .05$. **$p < .01$. ***$p < .001$.

Next, to assess whether P-TOI predicted the primary outcomes in the study above and beyond a general impression of instructors' level of approachability, support, and helpfulness we evaluated the primary results were robust to including instructors' approachability as covariates.

We retained only those covariates that were significant in the primary models for this analysis (see Table 25). Tests for H1d and H2 remained non-significant, and the test for H1a was no longer significant. Tests for H1b and H1c hovered around the significance threshold of p = .05 (.054 and .04, respectively). Thus, when controlling for instructor approachability, P-TOI no longer predicts dropping or considering dropping the course or concerns about being evaluated negatively, and neither does instructor approachability. Both P-TOI and instructor approachability predict course engagement with identical regression estimates (which can be viewed as effect sizes since all variables are standardized). TA approachability does not predict any of the outcomes in these analyses.

Table 24

Primary Results Controlling for Instructor and TA Approachability

	H1a: Dropping/ Intentions to Drop		H1b: Evaluative Concern		H1c: Course Engagement		H1d: Course Engagement × At Risk		H2: Help-seeking	
	β	SE	β	SE	β	SE	β	SE	β	SE
Intercept	-3.08***	0.81	0.04	0.07	0.01	0.07	0.34	0.21	0.08	0.07
P-TOI (S)	-0.39	0.27	-0.18	0.09	0.17*	0.09	0.30	0.21	-0.00	0.09
Instructor Approachability	-0.15	0.25	-0.04	0.09	0.17*	0.08	0.18*	0.08	0.05	0.09
TA Approachability	-0.48	0.21	-0.08	0.08	0.10	0.08	0.08	0.08	0.14	0.08
$\text{Subject}_{\text{CHEM}}$	2.20*	0.86	—		—		—		—	
$\text{Subject}_{\text{MATH}}$	1.33	0.92	—		—		—		—	
$\text{Subject}_{\text{PHYS}}$	1.01	0.91	—		—		—		—	
ITOI	—		-0.13	0.08	—		—		—	
At Risk	—		—		—		-0.37	0.22	—	
PTOI × At Risk	—		—		—		-0.13	0.22	—	
Academic Difficulty	—		—		—		—		0.39***	0.08

Note. P-TOI = Perceived Theories of Intelligence; ITOI = Implicit Theories of Intelligence; TA = Teaching Assistant

*p < .05. **p < .01. ***p < .001.

Longitudinal Analyses

Next in the exploratory analyses, we took advantage of the repeated measures design to answer the following two questions: Do the primary results hold when we disaggregate the weekly data points using hierarchical models? And did students' sense of self-efficacy change across time points as a function of their P-TOI scores? Since the data includes multiple observations, denoted by i, for each student, denoted by j, it has a two-level hierarchical structure with time points nested within students; data points are fixed-occasion with equal spacing of occasions (one week). There is a third level of nesting since students are also nested within courses, but our sample (9 courses) was not large enough to include a random effect for course. Time-varying level-1 response variables included Course Self-Efficacy (cse_{ij}), Evaluative Concern ($eval_{ij}$), and Course Engagement ($ceng_{ij}$). Academic Difficulty ($acdiff_{ij}$) and Week (i.e., the week survey was collected; $week_{ij}$) were level-1 subject and occasion-specific (time-varying) covariates.

Time-invariant level-2 variables included gender ($female_j$), underrepresented racial minority status (urm_j), GPA (gpa_j), and students' implicit theories of intelligence ($itoi_j$). Separate models were used to assess growth curves for Course Self-Efficacy, Evaluative Concern, and Course Engagement; these models were identical save the response variable. Assessing growth curves within a multilevel modeling framework allowed us to look at changes in students' trajectories across four weeks and make use of all available data. Per the preregistration, we limited the longitudinal analyses to participants who (a) completed at least 2 out of the 4 weekly surveys, and (b) completed both the intake and exit surveys. (In other words, participants who completed *at least* 4 out of 6 surveys.) The longitudnal dataset included 155 participants and 577 observations.

Similar to the confirmatory analyses, all hypotheses were tested using nested model comparison, and the full model was compared to a model that did not include the term of interest. The aim of the longitudinal analysis was to assess whether students' perceptions of their STEM instructor's implicit theories about intelligence at the end of the semester retrospectively predict *changes* in their level of Course Self-Efficacy, Course Engagement, and Evaluative Concern during the initial few weeks of the semester.

Time-invariant Level 2 Covariates

The items assessing students' Implicit Theories of Intelligence (*itoi*) were analyzed using an item-response model (partial credit model; Masters, 1982), and the scaled scores (*EAP* values) were entered into the hierarchical linear models. Grade Point Average (*gpa*; M = 3.58, SD = 0.49, range = 1.1–4.0) was a self-reported numeric variable (high school GPA substituted college GPA for students in their first semester). Gender was dummy-coded as *female*, with a value of 1 for females and other/non-binary and the value 0 for males. Underrepresented racial minority (*urm*) status was coded as 0 for White and Asian students, and 1 for all others.

Time-varying Level 1 Covariates

Academic difficulty (*acdiff*) was measured using a single item ("This past week, did you struggle academically in this course?") and included as a dummy variable (1 = *yes*; 0 = *no*). Age (*age*) was used as a self-reported numeric covariate.

We first tested whether P-TOI predicted changes in Course Self-Efficacy during the initial weeks of the semester (random intercepts for students and random slopes for time, controlling for covariates), specifying the two-level model as follows:

$$se_{ij} = \beta_0 + \beta_1 ptoi_j + \beta_2 week_i + \beta_3 ptoi_j \times week_i + \beta_4 acdiff_{ij} + \boldsymbol{\gamma Z_j} + \zeta_{1j} + \zeta_{2j} week_i + \varepsilon_{ij}$$

$$Z_j = \{x_{1j} = itoi,\ x_{2j} = gpa,\ x_{3j} = female,\ x_{4j} = urm\};$$

$(\zeta_{1j} \mid X_{ij}, Z_j) \sim N(0, \psi 11)$;

$(\zeta_{2j} \mid X_{ij}, Z_j) \sim N(0, \psi 22)$;

$(\varepsilon_{ij} \mid X_{ij}, Z_j) \sim N(0,\theta)$

Course Self-Efficacy for occasion *i* for student *j* was modeled as a function of variables represented by student-specific intercept ($\beta_0 + \zeta_{1j}$), student-specific slope ($\beta_1 + \beta_2 + \beta_3 + \zeta_{2j}$), and relevant level-1 and level-2 covariates. The β_4 coefficient corresponds to the time-varying items that assessed whether the student faced academic difficulty in each week. The $\mathbf{Z}_j$ vector corresponds to time-invariant student characteristics (gender, URM status, GPA, and ITOI). Given covariates, the random intercepts ζ_{1j} and random slopes ζ_{2j} have normal distributions, assumed to have zero mean, and variance of intercepts $\psi 11$, variance of random slopes for time ψ_{22}, and covariance of slopes and intercepts ψ_{12}. ε_{ij} is the level 1 residual term, also assumed to be normally distributed with mean 0 and variance θ, given the covariates and random effects.

We evaluated two other models that were identical to the model presented above, with Course Engagement and Evaluative Concern as the outcome variables. Since we were interested in change over time, the parameters of interest were the cross-level interactions between P-TOI and the time variable 'Week' (β_3). Table 25 presents the estimates for the hierarchical models predicting changes in Course Self-Efficacy, Course Engagement, and Evaluative Concern from P-TOI and covariates. The estimated ICCs for the three models range from 0.59 to 0.82, indicating that a sizable amount of total variability in the outcomes could be explained at the student level (which is also reflected in the variance estimates for intercepts and slopes, the former being relatively larger).

Model 1 (Course Self-Efficacy) indicated that each unit increase in P-TOI was associated with an estimated mean increase of 0.53 units in Course Self-Efficacy, controlling for other covariates (p = .006). On the other hand, facing academic difficulty was associated with an estimated *decrease* of 0.49 units in Course Self-Efficacy on average (p < .001), and being female was associated with an average 0.96 units lower Course Self-Efficacy, controlling for other covariates (p = 0.03). More importantly, the parameter of interest—the interaction between P-TOI and Week—was not significant. (For each unit increase in P-TOI per unit increase in week, self-efficacy was estimated to decrease, on average, by 0.01 units, p = 0.75). Thus, P-TOI did not retroactively predict change over time in students' self-efficacy. Similarly, Model 2 indicated that for each unit increase in P-TOI per unit increase in week, Course Engagement was estimated to increase, on average, by 0.26 units (p = .002). Course Engagement was estimated to decrease 0.08 units for each additional week (p = .003) and decrease 0.30 units for each additional unit of Academic Difficulty (p < .001), controlling for other covariates. No other coefficients in Model 2 were significant at the 5% alpha level (except the intercept). Thus, P-TOI did not retroactively predict a change over time in Course Engagement.

Model 3 estimated that each additional unit of P-TOI was associated with a mean decrease in Evaluative Concern by 0.32 units, controlling for all covariates; this effect was significant (p = .009). Further, each unit increase in academic difficulty was associated with an average increase of 0.20 units of Evaluative Concern (p = .04). The parameter that tests the hypothesis (P-TOI × Week) was not significant (each unit increase in P-TOI per unit increase in week, controlling for all covariates, was estimated to lead to a 0.02 unit decrease in Evaluative Concern). Thus, P-TOI did not retroactively predict changes in students' self-efficacy, course engagement, or evaluative concern across the initial weeks of the semester. However, controlling for covariates, P-TOI is associated positively with Self-Efficacy and Course Engagement, and negatively with Evaluative Concern.

Table 25

Multilevel Regression Estimates for Study 3

	Model 1: Self-Efficacy		Model 2: Course Engagement		Model 3: Evaluative Concern	
	β	*SE*	*β*	*SE*	*β*	*SE*
Fixed effects						
Intercept	-2.04	2.08	3.07***	0.69	2.19	1.50
Within-level						
Week	-0.00	0.05	-0.09**	0.03	-0.02	0.04
Academic Difficulty	-0.49***	0.14	-0.30***	0.08	0.20*	0.10
Between-level						
P-TOI	0.53**	0.19	0.26**	0.08	-0.33**	0.12
ITOI	0.12	0.07	0.01	0.02	-0.06	0.05
GPA	0.94	0.54	0.05	0.18	-0.63	0.39
Female	-0.96*	0.45	-0.13	0.15	0.00	0.33
URM	-0.84	0.44	-0.05	0.15	0.39	0.32
Cross-level						
P-TOI $\times$ Week	-0.0	0.05	-0.03	0.03	-0.02	0.04
Random effects						
ψ_{11}	4.34		0.50		1.77	
ψ_{22}	0.11		0.04		0.11	
ψ_{12}	0.97		0.37		0.44	
ρ	0.82		0.59		0.81	

Note. P-TOI = Perceived Theories of Intelligence; ITOI = Implicit Theories of Intelligence; GPA = Grade Point Average; URM = Underrepresented Minority. Within-level refers to variables that vary across time points, and between-level refers to variables that are invariant over time.

**p* < .05. ***p* < .01. ****p* < .001.

Given that P-TOI predicted Course Engagement and Evaluative Concern in the longitudinal analyses, these results replicate, to an extent, the findings from the primary analyses, despite the increased error variance induced by the random effects. This analysis also indicated that P-TOI is correlated with students' Course Self-Efficacy, thus adding another piece to the psychological puzzle of students' experience in STEM courses. However, students' perceptions of their instructors' theories about intelligence do not predict *changes* in any of these outcomes during the first few weeks of the semester. An important caveat here is that P-TOI was measured at the end of the semester. In future studies with lower time-constraints, researchers should concurrently measure P-TOI and students' psychological experiences over time to draw more reliable inferences.

Grades

Since we acquired data on student grades in the course, we could test how theories of intelligence are related to student performance in these courses. In this sample, there is a negative correlation between students' personal theories of intelligence (ITOI) and their final grade, $r(193) = -.15$, CI [-.28, -.01], $p = .04$. In a regression model that includes P-TOI and GPA as predictors, we find that, controlling for students' GPA ($\beta = 0.52$, $SE = 0.07$, $p = <.001$), students' perceptions of their instructors' theories of intelligence are positively related to grade ($\beta = 0.15$, $SE = 0.06$, $p = .02$; $F(1, 190) = 30.3$, $p < .001$). Including students' baseline perceptions (P-TOI [G]) to the model does not change the results ($\beta_{GPA} = 0.58$, $SE = 0.07$, $p < .001$, $\beta_{P\text{-}TOI(S)} = 0.23$, $SE = 0.08$, $p = .003$, $\beta_{P\text{-}TOI(G)} = -0.15$, $SE = 0.08$, $p = .06$, $F(3, 162) = 28.2$, $p < .001$).

Cross-Cultural Differences

We also assessed whether students from Asian background differed from non-Asian students in their theories and perceived theories of intelligence, and whether these differences, if they existed, correlated with their help-seeking behavior. Our results corroborate previous work showing that Asian students have a slightly less malleable view of intelligence, although this effect is not statistically significant; $\beta = - 0.21$, $SE = 0.12$, $p = .07$. However, we do see a significant interaction between being Asian and immigration status in predicting students' ITOI, $\beta_{ASIAN} = 0.03$, $SE = 0.13$, $p = .85$, $\beta_{IMMIGRANT} = 0.12$, $SE = 0.26$, $p = .64$, $\beta_{ASIAN \times IMMIGRANT} = -0.67$, $SE = 0.30$, $p = .03$; $F(3, 299) = 5.27$, $p = .001$. Non-Asian non-immigrants ($M = 4.31$, $SD = 1.14$, $n = 119$), non-Asian immigrants ($M = 4.45$, $SD = 0.95$, $n = 17$), and Asian non-immigrants ($M = 4.34$, $SD = 1.14$, $n = 102$). But Asian immigrants report more fixed theories of intelligence ($M = 3.71$, $SD = 1.17$, $n = 65$). (We should note that this may in explain the negative relationship in our data between students' personal theories of intelligence and course grade.)

We do not observe a relationship between being Asian or Asian immigrant with P-TOI (neither general nor specific). Being Asian or Asian immigrant also does not moderate the relationship between P-TOI (S) and course engagement, evaluative concern, or dropping/considering dropping the course, and we do not see a systematic difference in help-seeking behavior. Asian students received, on average, higher grades in the course ($\beta = 0.34$, $SE = 0.14$, $p = .02$; $F(1, 192) = 5.47$, $p = .02$), and the relationship between P-TOI (S) and course grade holds when controlling for Asian heritage ($\beta_{PTOI} = 0.17$, $SE = 0.07$, $p = .02$; $\beta_{ASIAN} = 0.34$, $SE = 0.14$, $p = .02$; $F(2, 191) = 5.71$, p = .003).

Discussion

The final study in this series aimed to assess whether students' perceptions of instructors' theories about intelligence affect their engagement and help-seeking behavior in rigorous STEM courses. Results from the preregistered analyses indicate that students' perceptions of their instructors' theories about intelligence positively predict course engagement, negatively predict concerns about being evaluated negatively, and correlate with lower dropout or intentions to drop out. Importantly, these analyses control for students' own theories about intelligence and other relevant covariates. We also find that P-TOI predicts—above and beyond how approachable and supportive students found the instructor—other subjective course-related outcomes like how much students enjoyed the course, how challenging they found the course, how competitive they found the course environment, and how much they report feeling motivated by and learning from the course (compared to other courses they have taken). Crucially, however, one of the key predictions of the study was not supported. We do not find evidence to suggest that students' perceptions of their STEM instructors' theories about intelligence predict help-seeking behavior during the initial weeks of the semester. Students' level of help-seeking seems influenced primarily by whether they were struggling in the course and by their level of general help-seeking (assessed at the beginning of the course).

There are several alternative explanations for our results. Although we tried to address this concern, students' perceptions of their instructors' theories about intelligence may be influenced by how much they like the instructor (a version of the Halo Effect; Nisbett & Wilson, 1977). It could also be the case that these perceptions are part and parcel of effective teaching, making it difficult to disentangle teaching effectiveness from perceptions of implicit theories in actual classrooms. Given that perceived instructor approachability and supportiveness and P-TOI are highly correlated and instructor approachability soaks up some of the variance in explaining the outcomes of interest away from P-TOI indicates that there is considerable overlap between two. Although a real-world context lends our results external validity, parsing instructors' theories about intelligence from their teaching effectiveness require further experimental study in the lab.

We also addressed a potential concern that students' experiences could be influenced more by their interactions with the teaching assistants. In contrast to instructor approachability, teaching assistants' approachability did not predict how much students enjoyed the course and how challenging and competitive they found it, and even though they report finding the teaching assistants more approachable, supportive, and helpful on average, and likely interacted more closely with them. Thus, these findings suggest that students' perceptions of the course instructors better explain students' sociopsychological experiences in these courses than their perceptions of the teaching assistants. However, when we include instructor and TA approachability in the primary hypothesis testing models, P-TOI only predicts course engagement, and the other effects are no longer significant. This is further evidence of the overlap between P-TOI and instructor approachability and supportiveness.

Nevertheless, if robust to replication, our null results paint an optimistic picture for instructors and the implicit theories theoretical framework. We do not find evidence to suggest that perceptions of STEM instructors' theories about intelligence differentially impact students from demographic backgrounds under-represented in STEM. It is important to note, however, that most students designated underrepresented in STEM in our sample were females; we do not know whether these results would replicate for a racially/ethnically diverse sample. Also of note here is the result that first-generation college students were less likely to complete the study, which may have biased the results. Although our convenience sample limits the generalizability of the findings, it also provides insights into a high-stakes, competitive STEM environment. Thus, our result adds a data point to the literature that might help assuage a concern that otherwise looms large in STEM education.

Finally, a crucial result from a theoretical perspective is that we do not find an empirical link between students' implicit theories and their perceptions of their instructors' implicit theories about intelligence. As noted earlier, help-seeking is both an adaptive academic behavior considered the hallmark of a malleable view of intelligence and a strategy often recommended to improve a student's growth mindset (Yeager et al., 2019). If students' implicit theories (or perceptions of others' implicit theories) interacted with help-seeking behavior, such recommendations would fail to support those who could benefit from them the most. Thus, if replicated, our null effect provides valuable information that bolsters the theoretical underpinnings of this domain.

Limitations and Future Directions

One of the critical limitations of the study is that the sample comes from a highly competitive public university in the United States, making it difficult to generalize to the larger student population. A majority of our sample is female, which adds another challenge to generalizability. Unlike previous studies (Canning et al., 2019; Rattan et al., 2012, 2018), in our sample, underrepresented students do not differ from non-underrepresented with regards to critical psychological variables and outcomes. Given that the university at which this research was conducted is highly selective (6.9% acceptance rate for the year the data were collected[3]), it is possible that underrepresented students in our sample are not representative of STEM students broadly. Additionally, 'weeder' or 'gateway' courses are not a universal phenomenon, which makes students' psychological experience within these courses of limited direct application outside the American context.

Another limitation of the study concerns measurement and psychometrics. Given the sheer number of factors that play an essential role in determining whether a student seeks help, it is crucial to operationalize help-seeking appropriately and in a context-sensitive manner, which we likely failed to do (evidenced by the lower internal consistency of the help-seeking composite). Moreover, many constructs were measured using single items on Likert-type response scales, which is far from ideal. Future studies should pay special attention to measuring the primary variables with higher fidelity when replicating these results.

Another area of further study would be to assess whether students are seeking help informally (peer study groups, learning centers). This might be especially relevant for students from more collectivistic backgrounds, like Asian and Asian-American students. Although we did not find a difference between Asian/Asian-American students and non-Asian students in their perceptions of instructors' theories of intelligence or help-seeking seeking, we did replicate the finding that Asian students have a less malleable view of intelligence. Future studies can address whether this difference results from variations in students' definitions of intelligence (knowledge vs. innate ability) and on the attribution of academic success to ability vs. effort.

Future work should also consider whether instructors' demographic characteristics (like gender and race/ethnicity) or perceived personality traits affect how students perceive their theories about intelligence and try to isolate how the messages instructors give to students about their potential/abilities affect student outcomes apart from the ones we considered. Additionally, pinpointing specific cues or aspects of the course structure/syllabus that contribute to students' perceptions would be enlightening for practitioners as well as researchers.

[3] https://opa.berkeley.edu/campus-data/common-data-set

Conclusion

The current study extends the burgeoning literature on perceptions of significant others' theories about intelligence in academic contexts. Importantly, we also show that students' perceptions of an internal state of a STEM instructor is correlated with their engagement, dropout intentions, and several other important psychological factors in the course. In our sample, however, these perceptions were unrelated to academic help-seeking and did not predict different outcomes for students from different backgrounds. If replicated, these results can help fill an essential gap in the growth mindset literature.

General Discussion

A core tenet of social psychology is that neither the person nor the environment alone, but the interaction between the two that determines behavior: $B = f(P, E)$ (Lewin, 1951). A second core tenet states that situational influence on behavior depends not on the objective properties of the situation but on the individual's subjective *construal* (Ross & Nisbett, 1991). These principles are especially relevant for motivation research, as noted by Atkinson (1964): "The focal point of interest in the study of motivation is the task of constructing a useful theoretical conception of how different factors *combine*, at a particular time, to influence the direction, vigor, and persistence of an individual's behavior in a given situation" (p. viii). Using these principles, the work presented in this thesis examined a common thread related to students' *perceptions* of their instructors' theories about intelligence—can these perceptions be accurately and reliably measured, do they covary with students' construal of the academic environment, and are they associated with students' willingness to reach out for help when needed?

The first study (Chapter 2) addressed whether college students' perceptions of their instructors' beliefs about intelligence and students' attitude toward help-seeking were correlated. We found that, on average, students more likely to perceive instructors at their university to hold a malleable view of intelligence were also more likely to report positive attitudes toward help-seeking and lower concerns about perceived incompetence. The same students also experienced lower competition among peers, lower feelings of impostorism, and lower belonging uncertainty. After establishing an association between perceptions of instructors' theories of intelligence and attitude toward help-seeking, our objective in the subsequent study (Chapter 3) was to design an instrument for measuring perceived implicit theories of intelligence (P-TOI) using a measurement framework that integrates theoretical and empirical investigation. We sampled high school and adult students, and the instrument exhibited good psychometric properties and sufficient reliability when administered to this sample (with some caveats for validity, explained below).

The aim of the final study (Chapter 4) was to further validate the instrument and test the association between P-TOI and students' help-seeking behavior within rigorous STEM (Science, Technology, Engineering, and Mathematics) courses. Results bolstered the reliability and validity evidence for the P-TOI instrument. Additionally, we observed that students who perceived their STEM instructors to be incremental theorists were, on average, more likely to report higher course engagement and subjective motivation, lower perceived competitiveness, were less concerned about being evaluated negatively, and had lower odds of dropping or considering dropping the course. Although we controlled for multiple demographic and psychological variables, the observational nature of our data precludes us from making causal inferences.

Across the three studies, we observed that students' perceptions of their instructors' theories about intelligence correlated with *affective* and *cognitive* aspects of their attitude toward academic help-seeking. However, we consistently failed to find evidence of an association between these perceptions and help-seeking *behavior*. Students in our sample asked for help when needed, given that facing academic difficulty was the strongest (and often only) predictor of self-reported help-seeking behavior. On the other hand, help-seeking was unrelated to implicit theories of intelligence, personal or perceived. In Study 1, we found that students' self-reported help-seeking behavior was predicted primarily by their confidence in their ability to self-regulate their learning and whether they were struggling academically. In line with previous research

(Karabenick & Knapp, 1988; Newman, 2002; A. M. Ryan et al., 1998; Tessler & Schwartz, 1972), students with higher self-efficacy and lower academic difficulty reported seeking more help. Perceived theories of intelligence did, however, predict the extent to which students perceived help-seeking as less threatening. In other words, students who perceived instructors to hold a malleable view of intelligence were less concerned they would be viewed as incompetent if they asked for help. This result from Study 1 is replicated in Studies 2 and 3; students who viewed their instructors/teachers as holding malleable views of intelligence were less concerned about being evaluated negatively.

In the present studies, students' personal theories of intelligence were correlated moderately with their perceptions of instructors' theories of intelligence. More interestingly, students' perception of their instructors' implicit theories was a more robust predictor of factors related to students' psychological experience. In our samples, unlike previous work (Blackwell et al., 2007; Claro et al., 2016; Yeager et al., 2013, 2016), we did not find a relationship between students' theories of intelligence and course grades, but grades were positively related to perceptions of instructors' theories of intelligence (Study 3). We also found, consistently, that when students perceived their instructors to view intelligence as more malleable, they perceived the learning environment as less competitive. This finding may especially relevant for contexts inherently plagued by high levels of competition (e.g., STEM; Bian et al., 2018; Leslie et al., 2015; Meyer et al., 2015), as the competitiveness of a classroom environment results primarily from classroom goals and structures, which are determined by the teacher.

However, an important caveat for interpreting our results is that perceived implicit theories of intelligence, as measured in studies described in Chapters 3 and 4, might be partially confounded with perceived instructor support. Despite good reliability, a qualitative reanalysis of item content, as well as some empirical results, raised questions about the latent construct being measured. Three pieces of evidence point in this direction. First, in Study 2, item 'Encourage' in the original set for the P-TOI instrument, which inquired about students' perception of instructor supportiveness, correlated highly with all other items, which serves as an indication that the latent construct the P-TOI instrument is measuring may be slightly different than the one intended. Second, the correlations between the P-TOI instrument are high and positive when students answered the items for instructors they considered effective, likable instructors, and indistinguishable from zero for instructors they considered ineffective. And finally, in Study 3, when we included the items on instructor approachability into the models that test our confirmatory hypotheses, the significant effects diminished or became non-significant, indicating that instructor 'approachability, supportiveness, and helpfulness' explains some of the variance that the P-TOI measure would otherwise explain. Although it could be the case that perceptions of instructors' theories of intelligence and perceived instructor support are distinct but highly correlated in the real world, we cannot conclude that from the current studies. These indications present a major limitation of the current work, and we present other limitations and future directions below.

Limitations and Future Directions

A fundamental limitation of our studies is the use of convenience samples for all empirical investigations. It is possible, even likely, that the variation present in the general population is not reflected in our samples, which were recruited primarily from selective universities in North America. The composition of our sample biases our estimates and limits their interpretation and generalizability. Additionally, all data presented are correlational and thus, we cannot—and do not—make causal claims. There are many possibilities for potential confounding, and thus, reverse causation and self-selection are key concerns in the interpretation of these results. Although our studies were sufficiently powered, especially Study 1, given the correlational nature of the data and the small effects of some correlations, we must also be wary that variables could be connected through causal structures that are unrelated to the constructs of theoretical interest (i.e., the crud factor; Meehl, 1990; see also Orben & Lakens, 2020).

Another crucial limitation is the abundant use of single-item self-report measures, many of which created ad hoc. We cannot assess the reliability of these measures, and although face-valid, they are not linked meaningfully to a construct. As Study 3 exemplifies, measuring a construct well is a long, arduous process, one that may still fall short of achieving its goal. Here, we focused on measuring a single construct with high fidelity at the cost of others. In future studies, we will aim to pare down the number of constructs studies and focus on better measurement. Conducting experiments can aid in this goal and allow for causal interpretation of the data.

As we mentioned earlier, due to the more intensive interaction between students and teachers in K-12, teachers of younger students might be able to gauge their students' abilities accurately, and relatedly, students might also be more likely to pick on cues from the teachers in these contexts. Additionally, there might be stronger, more direct cues relevant for one's potential in a K-12 classroom, which may be especially relevant for students in their more formative years when students are, to reiterate Jussim (1990), still forming "self-conceptions of their academic skills" (p. 24). Thus, P-TOI construct might be better suited for younger students. This claim is supported by the finding in Study 2 (Chapter 3) that the P-TOI items show differential functioning when comparing high school and adult students (with all the caveats of the sample used). Future studies can test this conjecture by replicating, for example, Study 3 with a younger sample and behavioral (as opposed to self-reported) measures of academic help-seeking.

Thus, the line of research presented here also has the potential to resurrect a generative line of research on expectancy effects, which shows that teachers' expectations can affect student outcomes (Rosenthal, 1987, 1991; Rosenthal & Jacobson, 1966, 1968a, 1968b). Research in this domain has traditionally focused on the effect of teacher expectancies on student achievement and whether these expectancies reflect accurate predictions or biases that result in self-fulfilling prophecies (Jussim, 1990). Researchers have argued that when controlling for past performance and motivation, correlations between teacher expectations and student performance are reduced considerably (Jussim, 1989, 1990; Williams, 1976). Expectancy effects are posited to be influenced by three factors: attributes of the perceiver, attributes of the target, and situational factors. Jussim (1990) notes, additionally, that expectancy effects are likely mediated by "awareness of the perceiver's expectations" (p. 24). In line with this claim, we posit that within the interaction of all three factors might be an important variable: the perception of a target's (teacher) expectancies by a perceiver (student) in situations where the target's judgement plays a pivotal role (classroom). The current thesis investigated whether students form (measurable) impressions of teacher expectancies about their intellectual potential and whether these impressions influence their motivation. We have presented some preliminary evidence suggesting it may be so. Further research can hopefully shed light on these questions and help us understand moderating conditions under which this effect could be larger or smaller.

Knowing that a perceiver holds a particular view about oneself can allow one to confirm or counteract those perceptions based on one's goals. And in a classroom, the goals are often determined by the teachers. Do students care more about looking smart or learning? Are they more worried about their grades or about learning and understanding? Teachers play a pivotal role in determining the answers to these questions and in determining the psychological experience of their students, as our results corroborate.

Conclusions

In the current work, we find that although help-seeking behavior is not associated with students' perceptions of educators' views of intelligence, these perceptions are related to several psychological variables that have downstream consequences for student motivation. We find that students who perceive their instructors to hold a malleable view of intelligence view help-seeking as adaptive, feel less worried about negative evaluation, experience lower peer-to-peer competition, and report feeling more motivated, engaged, and edified.

An important insight from social psychology is the immense power of social context in shaping human behavior. Social learning theory (Bandura, 1986; 2001), one of the guiding theories for the current work, reminds us that humans are meant to learn from others. A classroom is, first and foremost, a social context—a powerful social environment that shapes our understanding of the world, gives us the tools to function in it, and provides us with skills for communicating effectively with others. Science classrooms can be even more important, for they inspire those who help us push our shared understanding a little further.

Standing the front of a classroom, burdened by the curse of knowledge, with only a vague memory of what it was like to be a student, it can be easy for teachers to forget that students pay close attention to what teachers say and do. The research presented here serves as a reminder that students' experience in a classroom is significantly influenced by their teachers, and they likely want from their teacher not only knowledge, but also acknowledgement of their ability and potential. Students ascribe immense weight to teachers' actions and utterances, and teachers' impressions can act as calipers through which students assess their own caliber. Thus, the study of perceived implicit theories may allow for valuable insight into the processes that facilitate motivation and learning.

References

Ackerman, P. L., Kanfer, R., & Beier, M. E. (2013). Trait complex, cognitive ability, and domain knowledge predictors of baccalaureate success, STEM persistence, and gender differences. *Journal of Educational Psychology, 105*(3), 911–927. https://doi.org/10.1037/a0032338

Adams, R. J., & Khoo, S. T. (1996). *ConQuest* [Computer software]. Australia Council for Educational Research.

Adams, R. J., Wu, M., & Wilson, M. (2012). *ConQuest 3.0* [Computer software]. Australia Council for Educational Research.

Ahlmann-Eltze, C., & Patil, I. (2021). *ggsignif: R package for displaying significance brackets for 'ggplot2'*. doi:10.31234/osf.io/7awm6

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research, 4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2

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

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

Ames, R., & Lau, S. (1982). An attributional analysis of student help-seeking in academic settings. *Journal of Educational Psychology, 74*(3), 414–423. https://doi.org/10.1037/0022-0663.74.3.414

Anvari, F., & Lakens, D. (2021). Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology, 96*, Article 104159. https://doi.org/10.1016/j.jesp.2021.104159

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113–125. https://doi.org/10.1006/jesp.2001.1491

Atkinson, J.W. (1964). *An introduction to motivation.* Van Nostrand.

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

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

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25*(5), 729–735. https://doi.org/10.1037/0012-1649.25.5.729

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

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Information Age Publishing.

Barger, M. M., Xiong, Y., & Ferster, A. E. (2022). Identifying false growth mindsets in adults and implications for mathematics motivation. *Contemporary Educational Psychology, 70*, Article 102079. https://doi.org/10.1016/j.cedpsych.2022.102079

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01.

Be, J., Broïdo, H., Lam, P. F., Lee, J. Ha. (n.d.). Examining the enrollment and dropout rates of URMs in STEM courses at UC Berkeley. Retrieved July 28, 2023, from http://datadiversity.berkeley.edu/Weeder_Courses_Final.pdf

Berry, J. W. (1984). Towards a universal psychology of cognitive competence. *International Journal of Psychology, 19*(1-4), 335–361. https://doi.org/10.1080/00207598408247536

Bian, L., Leslie, S.-J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist, 73*(9), 1139–1153. https://doi.org/10.1037/amp0000427

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246–263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Boaler, J. (2013). Ability and mathematics: The mindset revolution that is reshaping education. *FORUM, 55*(1), 143–152. https://doi.org/10.2304/forum.2013.55.1.143

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–459. https://doi.org/10.1007/BF02293801

Bohns, V. K., & Flynn, F. J. (2010). "Why didn't you just ask?" Underestimating the discomfort of help-seeking. *Journal of Experimental Social Psychology, 46*(2), 402–409. https://doi.org/10.1016/j.jesp.2009.12.015

Boring, E.G. (1923). Intelligence as the tests test it. *New Republic, 36*, 35–37.

Brainard, S. G., & Carlin, L. (1998). A six-year longitudinal study of undergraduate women in engineering and science. *Journal of Engineering Education, 87*(4), 369–375. https://doi.org/10.1002/j.2168-9830.1998.tb00367.x

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology, 47*(6), 1191–1205. https://doi.org/10.1037/0022-3514.47.6.1191

Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences, 109*(2), 425–430. https://doi.org/10.1073/pnas.1106077109

Broda, M., Yun, J., Schneider, B., Yeager, D. S., Walton, G. M., & Diemer, M. (2018). Reducing inequality in academic success for incoming college students: A randomized trial of growth mindset and belonging interventions. *Journal of Research on Educational Effectiveness, 11*(3), 317–338. https://doi.org/10.1080/19345747.2018.1429037

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322.

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. https://doi.org/10.1177/0956797619897588

Burnette, J. L., Billingsley, J., Banks, G. C., Knouse, L. E., Hoyt, C. L., Pollack, J. M., & Simon, S. (2023). A systematic review and meta-analysis of growth mindset interventions: For whom, how, and why might such interventions work? *Psychological Bulletin, 149*(3–4), 174–205. https://doi.org/10.1037/bul0000368

Butler, R. (1998). Determinants of help seeking: Relations between perceived reasons for classroom help-avoidance and help-seeking behaviors in an experimental context. *Journal of Educational Psychology, 90*(4), 630–643. https://doi.org/10.1037/0022-0663.90.4.630

Butler, R. (2006). An achievement goal perspective on student help seeking and teacher help giving in the classroom: Theory, research, and educational implications. In S. A. Karabenick, & R. S. Newman (Eds.), *Help seeking in academic settings: Goals, groups, and contexts* (pp. 15–44). Routledge.

Butler, R., & Neuman, O. (1995). Effects of task and ego achievement goals on help-seeking behaviors and attitudes. *Journal of Educational Psychology, 87*(2), 261–271. https://doi.org/10.1037/0022-0663.87.2.261

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), Article eaau4734. https://doi.org/10.1126/sciadv.aau4734

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0. https://CRAN.R-project.org/package=pwr

Chen, X. (2013). STEM Attrition: College Students' Paths into and out of STEM Fields: Statistical analysis report. *National Center for Education Statistics.* Retrieved from https://files.eric.ed.gov/fulltext/ED544470.pdf

Chestnut, E., Lei, R., Leslie, S.-J., & Cimpian, A. (2018). The myth that only brilliant people are good at math and its implications for diversity. *Education Sciences, 8*(2), 65. https://doi.org/10.3390/educsci8020065

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292. https://doi.org/10.3102/0013189X16656615

Clance, P. R. (1985). *Clance Impostor Phenomenon Scale (CIPS)* [Database record]. APA PsycTests. https://doi.org/10.1037/t11274-000

Clance, P. R., & Imes, S. A. (1978). The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, Research & Practice, 15*(3), 241–247. https://doi.org/10.1037/h0086006

Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences, 113*(31), 8664–8668. https://doi.org/10.1073/pnas.1608207113

Cocodia, E. A. (2014). Cultural perceptions of human intelligence. *Journal of Intelligence, 2*(4), 180–196. https://doi.org/10.3390/jintelligence2040180

Costa, A., & Faria, L. (2018). Implicit theories of intelligence and academic achievement: A meta-analytic review. *Frontiers in Psychology, 9*, Article 829. https://doi.org/10.3389/fpsyg.2018.00829

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

Das, J.P. (1994). Eastern views of intelligence. In R. J. Sternberg (Ed.), *Encyclopaedia of Human Intelligence* (Vol. 1, pp. 387–391). Macmillan.

Dawson-Tunik, T. L., Commons, M., Wilson, M., & Fischer, K. W. (2005). The shape of development. *European Journal of Developmental Psychology, 2*(2), 163–195. https://doi.org/10.1080/17405620544000011

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology, 53*(6), 1024–1037. https://doi.org/10.1037/0022-3514.53.6.1024

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dember, W. N. (1974). Motivation and the cognitive revolution. *American Psychologist, 29*(3), 161–168. https://doi.org/10.1037/h0035907

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

Diener, C. I., & Dweck, C. S. (1980). An analysis of learned helplessness: II. The processing of success. *Journal of Personality and Social Psychology, 39*(5), 940–952. https://doi.org/10.1037/0022-3514.39.5.940

Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Kim-Prieto, C., Choi, D. W., & Oishi, S. (2009). New measures of well-being. In E. Diener (Ed.), *Assessing well-being: The collected works of Ed Diener* (pp. 247–266). Springer. https://doi.org/10.1007/978-90-481-2354-4_12

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences, 108*(19), 7716–7720. https://doi.org/10.1073/pnas.1018601108

Duckworth, A. L., Quirk, A., Gallop, R., Hoyle, R. H., Kelly, D. R., & Matthews, M. D. (2019). Cognitive and noncognitive predictors of success. *Proceedings of the National Academy of Sciences*, *116*(47), 23499–23504. https://doi.org/10.1073/pnas.1910510116

Dutilh, G., Sarafoglou, A., & Wagenmakers, E. J. (2021). Flexible yet fair: Blinding analyses in experimental psychology. *Synthese, 198*(23), 5745–5772. https://doi.org/10.1007/s11229-019-02456-7

Dweck, C. S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology, 31*(4), 674–685. https://doi.org/10.1037/h0077149

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

Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*(1), 5–12.

Dweck, C. S. (2000). *Self-Theories: Their Role in Motivation, Personality, and Development.* Psychology Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success.* Random House.

Dweck, C. S. (2008). Mindsets and math/science achievement. Carnagie Corporation of New York, Institute for Advanced Study, Commission on Mathematics and Science Education.

Dweck, C. S. (2011). Implicit theories. In P. V. Lange, A. Kruglanski, & T. Higgins (Eds.), *Handbook of Theories in Social Psychology, 2*, 43–61.

Dweck, C. S. (2016). Growth mindset interventions yield impressive results. The Conversation. Retrieved from https://theconversation.com/growth-mindset-interventions-yield-impressive-results-97423

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

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

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science, 14*(3), 481–496. https://doi.org/10.1177/1745691618804166

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215–225. https://doi.org/10.1177/0146167295213003

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* Chapman & Hall/CRC Press.

Fisher, J. D., Nadler, A., & Whitcher-Alagna, S. (1982). Recipient reactions to aid. *Psychological Bulletin, 91*(1), 27–54. https://doi.org/10.1037/0033-2909.91.1.27

Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., & Beilock, S. L. (2017). The math anxiety-performance link: A global phenomenon. *Current Directions in Psychological Science, 26*(1), 52–58. https://doi.org/10.1177/0963721416672463

Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA.

Flury, J. M., & Ickes, W. (2007). Having a weak versus strong sense of self: The Sense of Self Scale (SOSS). *Self and Identity, 6*(4), 281–303. https://doi.org/10.1080/15298860601033208

Freud, S. (1915). Instincts and their vicissitudes. *Standard Edition, 14*(1957), 109–40.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., & Qian, J. (2021). *Package 'glmnet'* [Computer software]. https://cran.r-project.org/web/packages/glmnet/

Friedman, J., Tibshirani, R., Hastie, T. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22. doi:10.18637/jss.v033.i01.

Fuesting, M. A., Diekman, A. B., Boucher, K. L., Murphy, M. C., Manson, D. L., & Safer, B. L. (2019). Growing STEM: Perceived faculty mindset as an indicator of communal affordances in STEM. *Journal of Personality and Social Psychology, 117*(2), 260–281. https://doi.org/10.1037/pspa0000154

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and Emotion, 27*, 199–223. https://doi.org/10.1023/A:1025007614869

Gardner, H. (1983). *Frames of mind: A theory of multiple intelligences*. Basic Books.

Gardner, H. (1987). *The mind's new science: A history of the cognitive revolution*. Basic Books.

Gill, R., & Keats, D. M. (1980). Elements of intellectual competence: Judgments by Australian and Malay university students. *Journal of Cross-Cultural Psychology, 11*(2), 233–243. https://doi.org/10.1177/0022022180112007

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*(1), 13–23.

Gottfredson, L. S. (1997b). Why g matters: The complexity of everyday life. *Intelligence, 24*(1), 79–132. https://doi.org/10.1016/S0160-2896(97)90014-3

Good, T. L. (1987). Two decades of research on teacher expectations: Findings and future directions. *Journal of Teacher Education, 38*(4), 32–47. https://doi.org/10.1177/002248718703800406

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700.

Gouëdard, P. (2021). Sky's the limit: Growth mindset, students, and schools in PISA. *PISA 2018.* OECD Publishing.

Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review, 3*, 5–39. https://doi.org/10.1007/BF01323661

Graham, S., & Chen, X. (2020). Attribution Theories. In *Oxford Research Encyclopedia of Education*. https://doi.org/10.1093/acrefore/9780190264093.013.892

Grayson, A., Miller, H., & Clarke, D. D. (1998). Identifying barriers to help-seeking: A qualitative analysis of students' preparedness to seek help from tutors. *British Journal of Guidance and Counselling, 26*(2), 237–253. https://doi.org/10.1080/03069889808259704

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software, 40*(3), 1-25. https://www.jstatsoft.org/v40/i03/

Gunderson, E. A., Hamdan, N., Sorhagen, N. S., & D'Esterre, A. P. (2017). Who needs innate ability to succeed in math and literacy? Academic-domain-specific theories of intelligence about peers versus adults. *Developmental Psychology, 53*(6), 1188–1205. https://doi.org/10.1037/dev0000282

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*(2), 139–150. https://doi.org/10.2307/2086306

Harrison, E., Drake, T., & Ots, R. (2022). finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling. R package version 1.0.5. https://CRAN.R-project.org/package=finalfit

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open, 5*(4). https://doi.org/10.1177/2158244015621113

Hess, R. D., Chang, C., & McDevitt, T. M. (1987). Cultural variations in family beliefs about children's performance in mathematics: Comparisons among people's republic of China, Chinese-American, and Caucasian-American families. *Journal of Educational Psychology, 79*(2), 179–188. http://dx.doi.org/10.1037/0022-0663.79.2.179

Haimovitz, K., & Dweck, C. S. (2016). What predicts children's fixed and growth intelligence mind-sets? Not their parents' views of intelligence but their parents' views of failure. *Psychological Science, 27*(6), 859–869. https://doi.org/10.1177/0956797616639727

Henderson, V. L., & Dweck, C. S. (1990). Motivation and achievement. In S. S. Feldman, & G. R. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 308–329). Harvard University Press.

Heyman, G. D., & Dweck, C. S. (1992). Achievement goals and intrinsic motivation: Their relation and their role in adaptive motivation. *Motivation and Emotion, 16*(3), 231–247.

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. https://doi.org/10.1037/0022-3514.77.3.588

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory.* Appleton-Century.

James, W. (1981). *The principles of psychology (Vol. II)*. Harvard University Press. (Original work published 1890)

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. *Advances in Experimental Social Psychology, 2*, 219–266. https://doi.org/10.1016/S0065-2601(08)60107-0

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review, 93*, 429–445.

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57*, 469–480.

Jussim, L. (1990). Social reality and social problems: The role of expectancies. *Journal of Social Issues, 46*(2), 9–34. https://doi.org/10.1111/j.1540-4560.1990.tb01921.x

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. https://doi.org/10.1207/s15327957pspr0902_3

Karabenick, S. A., & Knapp, J. R. (1988). Help seeking and the need for academic assistance. *Journal of Educational Psychology, 80*(3), 406–408. https://doi.org/10.1037/0022-0663.80.3.406

Karabenick, S. A., & Knapp, J. R. (1991). Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of Educational Psychology, 83*(2), 221–230. https://doi.org/10.1037/0022-0663.83.2.221

Karabenick, S. A., & Sharma, R. (1994). Perceived teacher support of student questioning in the college classroom: Its relation to student characteristics and role in the classroom questioning process. *Journal of Educational Psychology, 86*(1), 90–103. https://doi.org/10.1037/0022-0663.86.1.90

Keats, D. (1982). Cultural bases of concepts of intelligence: A Chinese versus Australian comparison. In *Proceedings of second Asian workshop on child and adolescent development* (pp. 67–75). Behavioral Science Research Institute.

Kelley, H. H. (1967). *Attribution theory in social psychology*. In Nebraska symposium on motivation. University of Nebraska Press.

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual review of psychology*, *31*, 457–501. https://doi.org/10.1146/annurev.ps.31.020180.002325

Kennedy, T. J., & Odell, M. R. (2014). Engaging students in STEM education. *Science Education International*, *25*(3), 246–258.

Kroeper, K. M., Fried, A. C., & Murphy, M. C. (2022). Towards fostering growth mindset classrooms: Identifying teaching behaviors that signal instructors' fixed and growth mindsets beliefs to students. *Social Psychology of Education, 25*(2-3), 371–398. https://doi.org/10.1007/s11218-022-09689-4

Lefcourt, H. M. (1966). Internal versus external control of reinforcement: A review. *Psychological Bulletin, 65*(4), 206–220. https://doi.org/10.1037/h0023116

Leslie, S. -J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347*(6219), 262–265. https://doi.org/10.1126/science.1261375

Lewin, K. (1938). Will and needs. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 283–299). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-024

Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). Harpers.

Licht, B. G., & Dweck, C. S. (1984). Determinants of academic achievement: The interaction of children's achievement orientations with skill area. *Developmental Psychology, 20*(4), 628–636. https://doi.org/10.1037/0012-1649.20.4.628

Li, Y., & Bates, T. C. (2019). You can't change your basic ability, but you work at things, and that's how we get hard things done: Testing the role of growth mindset on response to setbacks, educational attainment, and cognitive ability. *Journal of Experimental Psychology: General, 148*(9), 1640–1655. https://doi.org/10.1037/xge0000669

Limeri, L. B., Carter, N. T., Choe, J., Harper, H. G., Martin, H. R., Benton, A., & Dolan, E. L. (2020a). Growing a growth mindset: Characterizing how and why undergraduate students' mindsets change. *International Journal of STEM Education, 7*(35), 1–19. https://doi.org/10.1186/s40594-020-00227-2

Limeri, L. B., Choe, J., Harper, H. G., Martin, H. R., Benton, A., & Dolan, E. L. (2020b). Knowledge or abilities? How undergraduates define intelligence. *CBE—Life Sciences Education, 19*(1), Article ar5. https://doi.org/10.1187/cbe.19-09-0169

Linacre, J. M. (2010). Transitional categories and usefully disordered thresholds. *Online Educational Research Journal, 1*(3), 1–10.

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18*(3), 302–318. https://doi.org/10.1177/0146167292183006

Lüdecke D (2022). *sjlabelled: Labelled Data Utility Functions*. R Package version 1.2.0. https://doi.org/10.5281/zenodo.1249215

Lüdecke D (2018). sjmisc: Data and Variable Transformation Functions. *Journal of Open Source Software, 3(*26), 754. https://doi.org/10.21105/joss.00754

Lüdecke, D. (2023). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.14. https://CRAN.R-project.org/package=sjPlot

MacCoun, R., & Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature, 526*(7572), 187–189. https://doi.org/10.1038/526187a

MacCoun, R. J., & Perlmutter, S. (2017). Blind analysis as a correction for confirmatory bias in physics and in psychology. In S. O. Lilienfeld, & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 295-322). John Wiley & Sons. https://doi.org/10.1002/9781119095910.ch15

Mackintosh, N. J. (2011). History of theories and measurement of intelligence. In R. J. Sternberg, & S. B. Kaufman (Eds.), *The Cambridge Handbook of Intelligence* (pp. 3–19). Cambridge University Press.

Macnamara, B. N., & Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? A systematic review and meta-analysis with recommendations for best practices. *Psychological Bulletin, 149*(3-4), 133–173. https://doi.org/10.1037/bul0000352

Mapuranga, R., Dorans, N. J., & Middleton, K. (2008). A review of recent developments in differential item functioning (ETS Research Rep. No. 08-43). Princeton, NJ: Educational Testing Service.

Mari, L., Wilson, M., & Maul, A. (2023). *Measurement across the sciences: Developing a shared concept system for measurement.* Springer. https://doi.org/10.1007/978-3-031-22448-5

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Masten, A., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist, 53*(2), 205–220. https://doi.org/10.1037/0003-066X.53.2.205

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

Masters, G. N., & Wright, B. D. (1981). A model for partial credit scoring. Chicago, IL, USA: Statistical Laboratory, Department of Education, University of Chicago.

Masur, P. K. & Scharkow, M. (2020). *specr: Conducting and Visualizing Specification Curve Analyses*. https://CRAN.R-project.org/package=specr.

McClelland, D. C. (1985). How motives, skills, and values determine what people do. *American Psychologist*, *40*(7), 812–825. https://doi.org/10.1037/0003-066X.40.7.812

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). Toward a theory of motivation. In D. C. McClelland, J. W. Atkinson, R. A. Clark, & E. L. Lowell (Eds.), *The achievement motive* (pp. 6–96). Appleton-Century-Crofts. https://doi.org/10.1037/11144-002

McCutchen, K. L., Jones, M. H., Carbonneau, K. J., & Mueller, C. E. (2016). Mindset and standardized testing over time. *Learning and Individual Differences, 45*, 208–213. https://doi.org/10.1016/j.lindif.2015.11.027

McKeachie, W. J., Pintrich, P. R., & Lin, Y.-g. (1985). Teaching learning strategies. *Educational Psychologist*, *20*(3), 153–160. https://doi.org/10.1207/s15326985ep2003_5

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57*, 487–503. https://doi.org/10.1146/annurev.psych.56.091103.070258

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66*(1), 195–244. https://doi.org/10.2466/pr0.1990.66.1.195

Mehta, S., & Bunge, S. A. (2021). Socio-psychological experiences in gateway STEM courses. [Data set]. https://osf.io/mb7ks/

Mervis, J. (2011). Weed-out courses hamper diversity. *Science, 334*(6061), 1333. https://doi.org/10.1126/science.334.6061.1333

Meyer, M., Cimpian, A., & Leslie, S.J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology, 6,* Article 235. http://dx.doi.org/10.3389/fpsyg.2015.00235

Micari, M., & Calkins, S. (2021). Is it OK to ask? The impact of instructor openness to questions on student help-seeking and academic outcomes. *Active Learning in Higher Education, 22*(2), 143–157. https://doi.org/10.1177/1469787419846620

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences, 7*(3), 141–144. https://doi.org/10.1016/S1364-6613(03)00029-9

Miller, N. E. (1944). Experimental studies of conflict. In J. M. Hunt (Ed.), *Personality and the Behavior Disorders* (pp. 431–465). Ronald Press.

Miller, N. E. (1951). Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology* (pp. 435–472). Wiley.

Miller, N. E. (1959). Liberalization of basic SR concepts: Extensions to conflict behavior, motivation and social learning. *Psychology: A study of a science, Study, 1*, 196–292.

Moser, J. S., Schroder, H. S., Heeter, C., Moran, T. P., & Lee, Y. H. (2011). Mind your errors: Evidence for a neural mechanism linking growth mind-set to adaptive posterror adjustments. *Psychological Science, 22*(12), 1484–1489. https://doi.org/10.1177/0956797611419520

Muenks, K., Canning, E. A., LaCosse, J., Green, D. J., Zirkel, S., Garcia, J. A., & Murphy, M. C. (2020). Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class. *Journal of Experimental Psychology: General, 149*(11), 2119–2144. https://doi.org/10.1037/xge0000763

Muenks, K., Yan, V. X., & Telang, N. K. (2021). Who is part of the "mindset context"? The unique roles of perceived professor and peer mindsets in undergraduate engineering students' motivation and belonging. *Frontiers in Education, 6*, Article 633570. https://doi.org/10.3389/feduc.2021.633570

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. https://doi.org/10.1177/0146167209347380

Nadler, A. (1997). Personality and help seeking: Autonomous versus dependent seeking of help. In G. R. Pierce, B. Lakey, I. G. Sarason, B. R. Sarason (Eds.), *Sourcebook of social support and personality* (pp. 379–407). Springer. https://doi.org/10.1007/978-1-4899-1843-7_17

Nadler, A., & Fisher, J. D. (1986). The role of threat to self-esteem and perceived control in recipient reaction to help: Theory development and empirical validation. In *Advances in experimental social psychology* (pp. 81–122). Academic Press.

National Research Council. (1996). *National science education standards.* National Academies Press.

Nelson-Le Gall, S., & Glor-Scheib, S. (1985). Help seeking in elementary classrooms: An observational study. *Contemporary Educational Psychology, 10*(1), 58–71. https://doi.org/10.1016/0361-476X(85)90006-2

Neuwirth, E. (2022). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-3, https://CRAN.R-project.org/package=RColorBrewer.

Newman, R. S. (1990). Children's help-seeking in the classroom: The role of motivational factors and attitudes. *Journal of Educational Psychology, 82*(1), 71–80. https://doi.org/10.1037/0022-0663.82.1.71

Newman, R. S. (1994). Adaptive help seeking: A strategy of self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 283–301). Lawrence Erlbaum.

Newman, R. S. (2002). How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. *Theory into Practice, 41*(2), 132–138.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35*(4), 250–256. https://doi.org/10.1037/0022-3514.35.4.250

Okagaki, L., & Sternberg, R. J. (1993). Parental beliefs and children's school performance. *Child Development, 64*(1), 36–56. https://doi.org/10.1111/j.1467-8624.1993.tb02894.x

Orben, A., & Lakens, D. (2020). Crud (re) defined. *Advances in Methods and Practices in Psychological Science, 3*(2), 238–247. https://doi.org/10.1177/2515245920917961

Paek, I., & Wilson, M. (2011). Formulating the Rasch differential item functioning model under the marginal maximum likelihood estimation context and its comparison with Mantel–Haenszel procedure in short test and small sample conditions. *Educational and Psychological Measurement, 71*(6), 1023–1046. https://doi.org/10.1177/0013164411400734

Pang, W., Esping, A., & Plucker, J. A. (2017). Confucian Conceptions of Human Intelligence. *Review of General Psychology, 21*(2), 161–169. https://doi.org/10.1037/gpr0000103

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-set interventions are a scalable treatment for academic underachievement. *Psychological Science, 26*(6), 784–793. https://doi.org/10.1177/0956797615571017

Penner, A. M. (2015). Gender inequality in science. *Science, 347*(6219), 234–235. https://doi.org/10.1126/science.aaa3781

Pew Research Center. (2015). Public and scientists' views on science and society. Retrieved from https://www.pewresearch.org/science/2015/01/29/public-and-scientists-views-on-science-and-society/

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Princiotta, D., Goldstein, S. (2015). Intelligence as a conceptual construct: The philosophy of Plato and Pascal. In S. Goldstein, D. Princiotta, & J. Naglieri (Eds.), *Handbook of Intelligence: Evolutionary Theory, Historical Perspective, and Current Concepts*. Springer. https://doi.org/10.1007/978-1-4939-1562-0_6

R Core Team. (2023). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing [Computer software]. https://www.R-project.org/

Rattan, A., Good, C., & Dweck, C. S. (2012). “It's ok—Not everyone can be good at math”: Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*(3), 731–737. https://doi.org/10.1016/j.jesp.2011.12.012

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. https://doi.org/10.1037/pspi0000130

Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. R package version 2.3.6. https://CRAN.R-project.org/package=psych

Ritchie, S. J. (2015). *Intelligence: All that matters.* John Murray Learning.

Ritchie, S. J., Bates, T. C., Der, G., Starr, J. M., & Deary, I. J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging, 28*(2), 515–521. https://doi.org/10.1037/a0030820

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science, 29*(8), 1358–1369. https://doi.org/10.1177/0956797618774253

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. https://doi.org/10.1080/15298860290106805

Robitzsch, A., Kiefer, T., & Wu, M. (2022). *TAM: Test Analysis Modules*. R package version 4.1-4. https://CRAN.R-project.org/package=TAM

Rosen, S. (1983). Perceived inadequacy and help-seeking. In B. M. DePaulo, A. Nadler, & J. D. Fisher (Eds.), *New directions in helping (Vol. 2): Help-Seeking* (pp. 73-107). Academic Press.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Sage.

Rosenthal, R. (1985). From unconscious experimenter bias to teacher expectancy effects. In J. Dusek (Ed.), *Teacher expectancies* (pp. 37–65). Erlbaum.

Rosenthal, R. (1987). Pygmalion effects: Existence, magnitude, and social importance. *Educational Researcher, 16*(9), 37–40.

Rosenthal, R. (1991). Teacher expectancy effects: A brief update 25 years after the Pygmalion experiment. *Journal of Research in Education, 1*(1), 3–12.

Rosenthal, R., & Jacobson, L. F. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological Reports*, *19*(1), 115–118. https://doi.org/10.2466/pr0.1966.19.1.115

Rosenthal, R., & Jacobson, L. F. (1968a). Pygmalion in the classroom. *The Urban Review, 3*(1), 16–20. https://doi.org/10.1007/BF02322211

Rosenthal, R., & Jacobson, L. F. (1968b). Teacher expectations for the disadvantaged. *Scientific American, 218*(4), 19–23. https://www.jstor.org/stable/24926197

Rosenthal, R., & Rubie-Davies, C. M. (2015). How I spent my last 50-year vacation: Bob Rosenthal's lifetime of research into interpersonal expectancy effects. In C. M. Rubie-Davies, J. M. Stephens, & P. Watson (Eds.), In *The Routledge International Handbook of Social Psychology of the Classroom* (pp. 285–295). Taylor & Francis.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences, 1*(3), 377–386. https://doi.org/10.1017/S0140525X00075506

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Mcgraw-Hill Book Company.

Rotter, J. B. (1954). *Social learning and clinical psychology*. Prentice-Hall.

Rotter, J. B. (1975). Some problems and misconceptions related to the construct of internal versus external control of reinforcement. *Journal of Consulting and Clinical Psychology, 43*(1), 56–67. https://doi.org/10.1037/h0076301

Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high-and low-expectation teachers. *British Journal of Educational Psychology, 77*(2), 289–306. https://doi.org/10.1348/000709906X101601

Ryan, A. M., Gheen, M. H., & Midgley, C. (1998). Why do some students avoid asking for help? An examination of the interplay among students' academic efficacy, teachers' social–emotional role, and the classroom goal structure. *Journal of Educational Psychology, 90*(3), 528–535. https://doi.org/10.1037/0022-0663.90.3.528

Ryan, A. M., & Pintrich, P. R. (1997). "Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology, 89*(2), 329–341. https://doi.org/10.1037/0022-0663.89.2.329

Ryan, A. M., Pintrich, P. R., & Midgley, C. (2001). Avoiding seeking help in the classroom: Who and why? *Educational Psychology Review, 13*(2), 93–114. https://doi.org/10.1023/A:1009013420053

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Schmader, T., & Sedikides, C. (2018). State authenticity as fit to environment: The implications of social identity for fit, authenticity, and self-segregation. *Personality and Social Psychology Review, 22*(3), 228–259. https://doi.org/10.1177/1088868317734080

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology, 60*, Article 101832. https://doi.org/10.1016/j.cedpsych.2019.101832

Seligman, M. E. (1968). Chronic fear produced by unpredictable electric shock. *Journal of Comparative and Physiological Psychology, 66*(2), 402–411. https://doi.org/10.1037/h0026355

Seligman, M. E. (1972). Learned helplessness. *Annual Review of Medicine, 23*(1), 407–412.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour, 4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

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

Skaalvik, S., & Skaalvik, E. M. (2005). Self-concept, motivational orientation, and help-seeking behavior in mathematics: A study of adults returning to high school. *Social Psychology of Education, 8*(3), 285–302. https://doi.org/10.1007/s11218-005-3276-3

Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology, 15*(2), 201–293. https://doi.org/10.2307/1412107

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*(3), 171–195.

Stephens, N. M., Townsend, S. S., Markus, H. R., & Phillips, L. T. (2012). A cultural mismatch: Independent cultural norms produce greater increases in cortisol and more negative emotions among first-generation college students. *Journal of Experimental Social Psychology, 48*(6), 1389–1393. https://doi.org/10.1016/j.jesp.2012.07.008

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology, 49*(3), 607–627. https://doi.org/10.1037/0022-3514.49.3.607

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence.* Cambridge University Press.

Stevenson, H. W., Chen, C. & Lee, S. Y. (1993). Mathematics achievement of Chinese, Japanese, and American children: Ten years later. *Science, 259*(5091), 53–58. https://doi.org/10.1126/science.8418494

Stevenson, H., & Stigler, J. W. (1994). *Learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. Simon & Schuster.

Sun, X., Nancekivell, S., Gelman, S. A., & Shah, P. (2021). Growth mindset and academic outcomes: A comparison of US and Chinese students. *NPJ Science of Learning, 6*(1), 21. https://doi.org/10.1038/s41539-021-00100-z

Super, C. M. (1983). Cultural variation in the meaning and uses of children's "intelligence." In J. B. Deregowski, R. C. Annis, & S. Dziurawiec (Eds.), *Expiscations in Cross-cultural Psychology,* 199–212.

Tessler, R. C., & Schwartz, S. H. (1972). Help seeking, self-esteem, and achievement motivation: An attributional analysis. *Journal of Personality and Social Psychology, 21*(3), 318–326. https://doi.org/10.1037/h0032321

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tibshirani, R., Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J., & Tibshirani, R. J. (2012). Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *74*(2), 245–266. https://doi.org/10.1111/j.1467-9868.2011.01004.x

Tibshirani, R., Tibshirani, R., Taylor, J., Loftus, J., Reid, S., & Markovic, J. (2019). *selectiveInference: Tools for post-selection inference*. *R package version 1.2.5* [Computer software]. https://CRAN.R-project.org/package=selectiveInference

Tipton, E., Bryan, C., Murray, J., McDaniel, M. A., Schneider, B., & Yeager, D. S. (2023). Why meta-analyses of growth mindset and other interventions should follow best practices for examining heterogeneity: Commentary on Macnamara and Burgoyne (2023) and Burnette et al. (2023). *Psychological Bulletin, 149*(3-4), 229–241. https://doi.org/10.1037/bul0000384

Thibaut, J. W., & Riecken, H. W. (1955). Some determinants and consequences of the perception of social causality. *Journal of Personality*, *24*(2), 113–133. https://doi.org/10.1111/j.1467-6494.1955.tb01178.x

Thorndike, E. (1900). Mental fatigue. I. *Psychological Review, 7*(6), 547–579. https://doi.org/10.1037/h0069511

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review, 34(4), 273–286. https://doi.org/10.1037/h0070288

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*(4), 529–554. https://doi.org/10.1086/214483

Tolman, E. C. (1922). Can instincts be given up in psychology? *The Journal of Abnormal Psychology and Social Psychology, 17*(2), 139–152. https://doi.org/10.1037/h0064419

Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review, 45*(1), 1–41. https://doi.org/10.1037/h0062733

Tolstoy, L. (2002). *Anna Karenina.* Penguin Books. (Original work published 1877)

Torres Irribarra, D. & Freund, R. (2014). *Wright Map: IRT item-person map with ConQuest integration* [Computer software]. https://github.com/david-ti/wrightmap

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

Walton, G. M., & Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science, 29*(3), 219–226. https://doi.org/10.1177/0963721420904453

Weiner, B. (1972). Attribution theory, achievement motivation, and the educational process. *Review of Educational Research, 42*(2), 203–215. https://doi.org/10.3102/00346543042002203

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*(1), 3–25. https://doi.org/10.1037/0022-0663.71.1.3

Weiner, B. (1980). *Attribution theory and human motivation*. Holt, Reinhart & Winston.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548

Weiner, B. (1986). Attribution, emotion, and action. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 281–312). Guilford Press.

Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology, 82*(4), 616–622. https://doi.org/10.1037/0022-0663.82.4.616

Weiner, B. (2018). The legacy of an attribution approach to motivation and emotion: A no-crisis zone. *Motivation Science, 4*(1), 4–14. https://doi.org/10.1037/mot0000082

Weiner, B., & Kukla, A. (1970). An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology, 15*(1), 1–20. https://doi.org/10.1037/h0029211

Wentzel, K. R., & Wigfield, A. (1998). Academic and social motivational influences on students' academic performance. *Educational Psychology Review, 10*(2), 155–175. https://doi.org/10.1023/A:1022137619834

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software, 40*(1), 1–29. https://www.jstatsoft.org/v40/i01/

Wickham, H. (2022). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.0. https://CRAN.R-project.org/package=stringr>.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. doi:10.21105/joss.01686.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.2. https://CRAN.R-project.org/package=dplyr

Wilke C (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 1.1.1. https://CRAN.R-project.org/package=cowplot

Williams, T. (1976). Teacher prophecies and the inheritance of inequality. *Sociology of Education*, 223–236. https://doi.org/10.2307/2112234

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Routledge.

Wilson, M. (2011). Some notes on the term: "Wright Map." *Rasch Measurement Transactions, 25*(3), 1331.

Wilson, M. (2023). *Constructing measures: An item response modeling approach.* Taylor & Francis.

Wilson, M., Bathia, S., Morell, L., Gochyyev, P., Koo, B. W., & Smith, R. (2022). Seeking a better balance between efficiency and interpretability: Comparing the Likert response format with the Guttman response format. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000462

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education,13*(2), 181–208. https://doi.org/10.1207/S15324818AME1302_4

Wilson, T. D., Damiani, M., & Shelton, N. (2002). Improving the academic performance of college students with brief attributional interventions. In J. Aronson (Ed.), *Improving Academic Achievement* (pp. 89–108). Academic Press. https://doi.org/10.1016/B978-012064455-1/50008-7

Yang, S. Y., & Sternberg, R. J. (1997). Taiwanese Chinese people's conceptions of intelligence. *Intelligence*, *25*(1), 21–36. https://doi.org/10.1016/S0160-2896(97)90005-2

Yeager, D. S., Carroll, J. M., Buontempo, J., Cimpian, A., Woody, S., Crosnoe, R., Muller, C., Murray, J., Mhatre, P., Kersting, N., Hulleman, C., Kudym, M., Murphy, M., Duckworth, A. L., Walton, G. M., & Dweck, C. S. (2022). Teacher mindsets help explain where a growth-mindset intervention does and doesn't work. *Psychological Science, 33*(1), 18–32. https://doi.org/10.1177/09567976211028984

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., ... & Dweck, C. S. (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. https://doi.org/10.1037/edu0000098

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

Yeager, D. S., Walton, G. M., & Cohen, G. L. (2013). Addressing achievement gaps with psychological interventions. *Phi Delta Kappan, 94*(5), 62–65.

Yousaf, A. (2023). Relationship between mindset and intelligence. *Canadian Journal for New Scholars in Education, 14*(1), 87–97.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: the role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*(3), 663–676. https://doi.org/10.3102/00028312029003663

Zhang D (2022). *rsq: R-Squared and Related Measures*. R package version 2.5, https://CRAN.R-project.org/package=rsq.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix A

The R code to reproduce the analyses reported in this article is available via ResearchBox, at https://researchbox.org/870. The figures and statistical analyses were created using R Studio (Version 2023.06.1; RStudio, 2022) and R (Version 4.3.1, R Core Team, 2023) and R packages car (Version 3.1-2; Fox, 2019), cowplot (Version 1.1.1; Wilke, 2020), dplyr (Version 1.1.2; Wickham, 2023), finalfit (Version 1.0.6; Harrison et al., 2022), ggplot2 (Version 3.4.2; Wickham et al., 2009), ggrain (Version 0.0.3; Allen et al., 2021), ggsignif (Version 0.6.4; Ahlmann-Eltze & Patil, 2021), glmnet (Version 4.1-7; Friedman et al., 2010), lme4 (Version 1.1-34; Bates et al., 2015), lubridate (Version 1.9.2; Grolemund & Wickham, 2011), plyr (Version 1.8.8; Wickham, 2011), psych (Version 2.3.6; Revelle, 2023), pwr (Version 1.3-0; Champely, 2017), raincloudplots (Version 0.2.0; Allen et al., 2021), RColorBrewer (Version 1.1-3; Neuwirth, 2022), rsq (Version 2.5; Zhang, 2022), selectiveInference (Version 1.2.5; Tibshirani et al., 2019), sjlabelled (Version 1.2.0; Lüdecke, 2022), sjmisc (Version 2.8.9; Lüdecke, 2018), sjPlot (Version 2.8.14; Lüdecke, 2023), specr (Version 1.0.0; Masur & Scharkow, 2020), stringr (Version 1.5.0; Wickham, 2022), TAM (Version 4.1-4; Robitzsch et al., 2022), tidyverse (Version 2.0.0; Wickham et al., 2019), and WrightMap (Version 1.3; Torres Irribarra & Freund, 2014).

Appendix B

Supplementary Analyses for Study 1

Table B1

Regression Estimates from Linear Models predicting Help-Seeking with Demographic Variables

Variable	MSLQ	Help-seeking as Threatening	Help-seeking Perception	Self-Perception of Help	Help-Seeking Behavior	HS Composite
Age	-0.03 (0.03)	-0.09 (0.03)**	-0.06 (0.03)*	0.03 (0.03)	-0.16 (0.07)*	0.01 (0.03)
Female	0.03 (0.03)	0.05 (0.03)	0.19 (0.02)***	0.02 (0.03)	0.06 (0.04)	0.05 (0.03)
URM	-0.02 (0.03)	-0.05 (0.03)	0.10 (0.03)***	-0.01 (0.03)	0.00 (0.05)	0.02 (0.03)
First Generation	0.08 (0.03)**	0.01 (0.03)	-0.07 (0.03)*	0.03 (0.03)	0.04 (0.05)	0.05 (0.03)
Social Class	0.07 (0.03)*	-0.08 (0.03)**	-0.03 (0.03)	0.03 (0.03)	-0.03 (0.05)	0.07 (0.03)*
Semesters	-0.10 (0.03)***	-0.03 (0.03)	0.10 (0.03)***	-0.06 (0.03)*	-0.13 (0.05)**	-0.05 (0.03)
Transfer Status	-0.01 (0.03)	0.00 (0.03)	0.00 (0.03)	-0.02 (0.03)	-0.01 (0.05)	-0.01 (0.03)
GPA	0.09 (0.03)***	-0.15 (0.03)***	0.07 (0.03)**	0.18 (0.03)***	-0.08 (0.04)*	0.16 (0.03)***

Note. HS = Help-Seeking; MSLQ = Motivated Strategies for Learning Questionnaire; URM = Underrepresented Minority; GPA = Grade Point Average. The models do not include an intercept.
**p* < .05. ***p* < .01. ****p* < .001.

Primary Results with Exclusions

Below are the main results from Study 1 (Chapter 2) after limiting the data to participants who passed (a) two out of three attention checks (n = 1353) and (b) all attention checks (n = 857). The results remain unchanged, with one exception—among students who passed all attention checks, perceptions of competitiveness negatively predicted attitude toward help-seeking.

Two Out of Three Attention Checks

Consistent with the main results, perceiving instructors at one's institution to hold a malleable view of intelligence was correlated, on average, with more positive attitudes toward academic help-seeking, controlling for students' own mindset, and the experience of academic difficulty in the previous month, $\chi^2(1) = 47.5$, $p < .001$, $n = 1232$, $r_{partial} = .19$. (See Table B2 for standardized regression estimates and standard errors.) Testing the same model with impostor phenomenon as the dependent variable, results show that faculty growth mindset negatively predicted impostor feelings, $\chi^2(1) = 8.91$, $p = .003$, $n = 1232$, $r_{partial} = -.08$. Identical analyses with belonging uncertainty and perceptions of competitiveness as outcomes show that faculty growth mindset was associated with lower belonging uncertainty ($\chi^2(1) = 36.9$, $p < .001$, $n = 1232$, $r_{partial} = -.17$), as well as lower perceptions of competitiveness ($\chi^2(1) = 137$, $p < .001$, $n = 1232$, $r_{partial} = -.32$).

In multiple regression models that included students' growth mindset, GPA, and academic difficulty as covariates, perceptions of competitiveness did not predict attitude toward help-seeking, β = -0.04, *SE* = .03, $\chi^2(1)$ = 2.47, *p* = .12, *n* = 1268, $r_{partial}$ = -.03, but impostor phenomenon did, β = -0.31, *SE* = 0.03, $\chi^2(1)$ = 134, *p* < .001, *n* = 1268, $r_{partial}$ = -.31. Consistent with the main results, students' growth mindset was correlated moderately with perceived faculty growth mindset, *r*(1305) = .27, *p* < .001, CI [.22, .32], and students' higher in growth mindset were less likely to view help-seeking as threatening, *r*(1349) = -0.16, *p* < .001, CI [-.21, -.11].

Table B2

Standardized Regression Estimates for Main Results in Study 1 with Exclusions (n = 1353)

	Help-Seeking Composite		Impostor Phenomenon		Belonging Uncertainty		Competitiveness	
	β	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Faculty Growth Mindset	0.20***	0.03	-0.09**	0.03	-0.17***	0.03	-0.33***	0.03
Growth Mindset	0.14***	0.03	-0.12***	0.03	-0.04	0.03	-0.01	0.03
GPA	0.12***	0.03	0.05	0.03	-0.14***	0.03	0.00	0.03
Academic Difficulty (No)	0.08	0.05	-0.32***	0.05	-0.31***	0.05	-0.20***	0.05
Academic Difficulty (Yes)	-0.05	0.03	-0.15***	0.03	0.14***	0.03	0.09**	0.03
Adjusted R^2	.10		.07		.12		.13	

Note. GPA = Grade Point Average. Intercept terms have been removed.

**p* < .05. ***p* < .01. ****p* < .001.

All Attention Checks

When we subset the data to participants who passed all attention checks, the results are consistent with the main results and the results from the subset of participants who passed two out of three attention checks. Perceiving instructors at one's institution to hold a malleable view of intelligence was correlated, on average, with more positive attitudes toward academic help-seeking, controlling for students' own mindset, and the experience of academic difficulty in the previous month, $\chi^2(1)$ = 46.2, *p* < .001, *n* = 789, $r_{partial}$ = .23. (See Table B3 for standardized regression estimates and standard errors.)

Evaluating the same model with impostor phenomenon as the dependent variable, we find that faculty growth mindset negatively predicts students' impostor feelings, $\chi^2(1)$ = 18.1, *p* < .001, *n* = 789, $r_{partial}$ = -.08. Identical analyses with belonging uncertainty and competitiveness as dependent variables showed that faculty growth mindset was associated with lower belonging uncertainty ($\chi^2(1)$ = 35.1, *p* < .001, *n* = 789, $r_{partial}$ = -.20), as well as lower perceptions of competitiveness ($\chi^2(1)$ = 99.9, *p* < .001, *n* = 789, $r_{partial}$ = -.33).

In a multiple regressions model that included students' growth mindset, GPA, and academic difficulty as covariates, perceptions of competitiveness *did* predict attitude toward help-seeking, β = -0.07, *SE* = .04, $\chi^2(1)$ = 4.36, *p* = .04, *n* = 809, $r_{partial}$ = -.06. Additionally, students who reported higher impostor feelings also reported more negative attitude toward help-seeking, β = -0.31, *SE* = 0.03, $\chi^2(1)$ = 82, *p* < .001, *n* = 809, $r_{partial}$ = -.30. Students' growth mindset was once again correlated moderately with perceived faculty growth mindset, $r(831) = .28$, $p < .001$, CI [.22, .34], and students higher in growth mindset were less likely to view help-seeking as threatening, $r(809) = -.18$, $p < .001$ CI [-.24, -.11].

Table B3

Standardized Regression Estimates for Study 1 Primary Results with Exclusions (n = 857)

Variable	HS Composite		Impostor Phenomenon		Belonging Uncertainty		Competitiveness	
	β	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Faculty Growth Mindset	0.24***	0.04	-0.15***	0.04	-0.20***	0.03	-0.35***	0.03
Growth Mindset	0.13***	0.04	-0.10***	0.04	-0.04	0.03	-0.02	0.03
GPA	0.13***	0.04	0.07	0.04	-0.14***	0.04	0.01	0.04
Academic Difficulty (No)	0.09	0.06	-0.32***	0.06	-0.33***	0.06	-0.21***	0.06
Academic Difficulty (Yes)	-0.06	0.04	-0.16***	0.04	0.17***	0.04	0.10*	0.04
Adjusted R^2	.12		.08		.15		.14	

Note. HS = Help-Seeking; GPA = Grade Point Average. Intercept terms have been removed.

*$p < .05$. **$p < .01$. ***$p < .001$.

Appendix C

Supplementary Figures for Study 2 (Chapter 3)

Figure C1

Liked as Person Wright Map

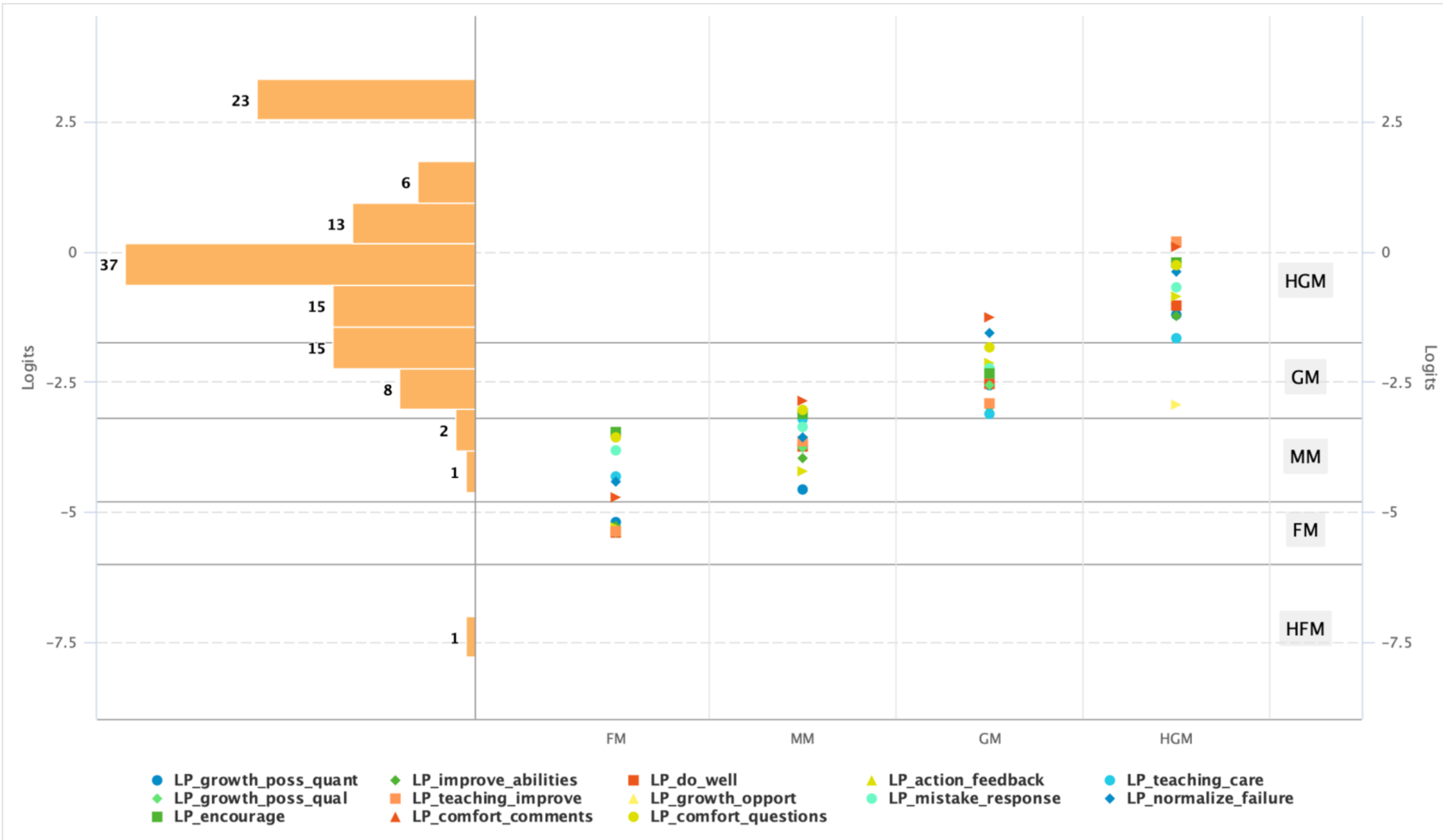
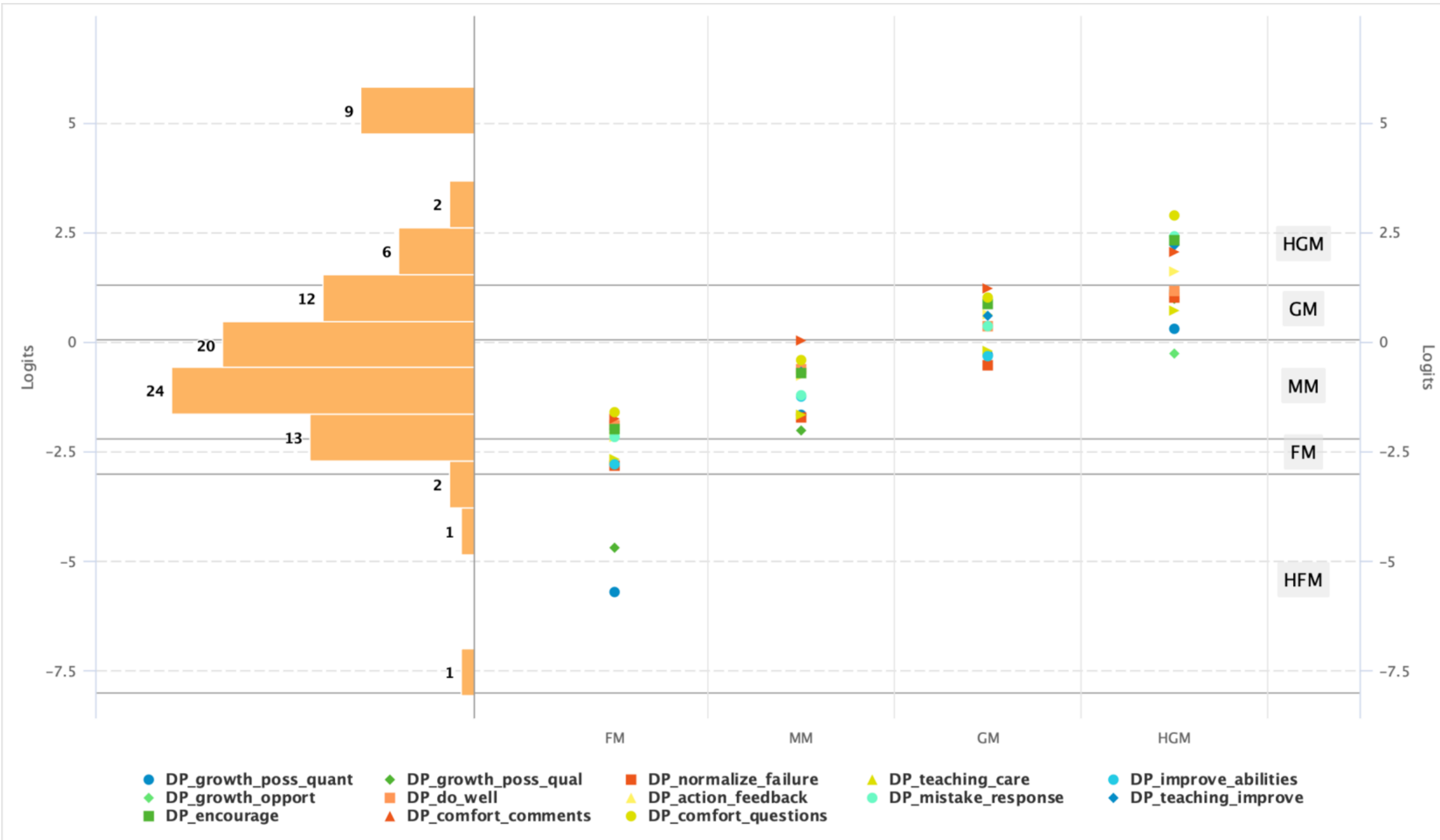


Figure C2

Disliked as Person Wright Map



Appendix D

Supplementary Analyses for Study 3

P-TOI Psychometric Properties

The data for both P-TOI (G) and P-TOI (S) fit the partial credit model well. There were 265 cases in the estimation for P-TOI (G) and 205 cases for P-TOI (S); both models estimated 30 parameters. EAP reliability was .81 for P-TOI (G) and .83 for P-TOI (S), and Cronbach's α was .82 for both. Only two items (Improve Abilities and Normalize Failure) have *MNSQ* values outside the 95% confidence interval, but in all three cases, the *MSNQ* are at the .75–1.33 boundary. Thus, item misfit is minimal and does not warrant concern.

Table D1

Item	P-TOI (G)						P-TOI (S)					
	δ	*SE*	*MNSQ*	*CI*	*t*	*p*	*δ*	*SE*	*MNSQ*	*CI*	*t*	*p*
Do Well	-0.52	0.07	1.01	[0.84, 1.16]	0.1	.92	-1.57	0.09	0.93	[0.80, 1.20]	-0.7	.48
Improve Abilities	-1.06	0.07	0.87	[0.83, 1.17]	-1.5	.13	**-1.91**	**0.10**	**0.74**	**[0.79, 1.21]**	**-2.6**	**.01**
Growth Possible	-1.81	0.08	0.86	[0.83, 1.17]	-1.7	.09	-2.27	0.11	0.93	[0.78, 1.22]	-0.6	.55
Teaching Care	-1.35	0.08	0.87	[0.82, 1.18]	-1.5	.13	-1.79	0.10	0.82	[0.77, 1.23]	-1.6	.11
Actionable Feedback	-0.51	0.08	1.13	[0.83, 1.17]	1.4	.16	-1.21	0.09	1.16	[0.80, 1.20]	1.6	.11
Growth Oppor.	-1.48	0.16	0.94	[0.83, 1.17]	-0.7	.48	-1.24	0.18	1.03	[0.83, 1.17]	0.4	.69
Mistake Response	-1.07	0.08	0.99	[0.83, 1.17]	-0.1	.92	-1.70	0.10	1.07	[0.79, 1.21]	0.7	.48
Normalize Failure	**-1.66**	**0.08**	**1.33**	**[0.82, 1.18]**	**3.3**	**.001**	**-2.01**	**0.11**	**1.35**	**[0.80, 1.20]**	**3.2**	**.002**

Note. P-TOI = Perceived Theories of Intelligence; (G) = General; (S) = Specific; *MSNQ* = Weighted Mean Square. Items with MNSQ values outside the 95% CIs are in bold.

Primary Results Controlling for P-TOI (G)

Below we present the results for the tests of primary results after including students' baseline scores on P-TOI (i.e., P-TOI (G)) as a covariate. The results do not change, but there is a reduction in effect sizes and the p-values, although significant at the α = .05 level, are larger.

(f) H1a stated that students who scored higher on Perceived Theories of Intelligence (P-TOI) instrument would be less likely to drop or consider dropping the course during the semester. After controlling for covariates, including P-TOI (G), each unit increase on P-TOI was associated with 50% lower odds of dropping or considering dropping the course, χ^2 (1) = 5.55, n = 147, p = 0.02 (*OR* = 0.50, CI [0.27, 0.89], $r_{partial}$ = -.70).

(g) Our second hypothesis, H1b, was that students who score higher on P-TOI would be less concerned about being evaluated negatively by instructors and peers. We found that for each unit increase on P-TOI, students were, on average, 0.19 units less concerned about being evaluated negatively, χ^2 (1) = 3.21, n = 145, p = .04 ($r_{partial}$ = -0.17).

(h) Next, we hypothesized that students who scored higher on P-TOI would self-report experiencing higher course engagement, controlling for the level of academic difficulty faced (H1c). Consistent with this hypothesis, we found that for each unit increase on P-TOI, students reported experiencing, on average, 0.20 units higher course engagement, χ^2 (1) = 3.41, n = 144, p = .03 ($r_{partial}$ = 0.18).

(i) We did not find evidence in support of H1d; that is, the interaction between P-TOI and at-risk status, based on demographic characteristics (URM status, gender, first-gen status), was not significant when predicting course engagement, χ^2(1) = 0.28, n = 144, p = 0.55 (β = -0.14, SE = 0.24, $r_{partial}$ = 0.05). That is, at-risk students (compared to not-at-risk students) did not differ in terms of course engagement if they perceived the instructor to hold a malleable view of intelligence.

(j) Further, scoring higher on P-TOI was not positively correlated with academic help-seeking, χ^2(1) = 0.05, n = 145, p = 0.81 (β = 0.03, SE = 0.10, $r_{partial}$ = 0.02). Thus, contrary to our H2 prediction, scores on P-TOI did not predict actual help-seeking behavior during the initial weeks of the course.

As in the primary analyses, the first three hypotheses were empirically supported: students who scored higher on P-TOI were less concerned about negative evaluation from instructors and classmates, more engaged, and less likely to drop or consider dropping the course. The results did not indicate that students at risk for attrition experienced higher course engagement as a function of P-TOI, or that P-TOI was associated with help-seeking behavior during the initial weeks of the semester.

Table D2

Regression Estimates from Primary Analysis in Study 3, controlling for P-TOI (G)

	H1a: Dropping		H1b: Evaluative Concern		H1c: Course Engagement		H1d: Course Engagement × At Risk		H2: Help-seeking	
	β	SE	β	SE	β	SE	β	SE	β	SE
Intercept	-2.90*	1.19	-0.12	0.32	0.36	0.31	0.40	0.33	-0.16	0.34
P-TOI (S)	-0.69*	0.30	-0.19*	0.10	0.20*	0.09	0.33	0.23	0.03	0.10
P-TOI (G)	-0.01	0.32	-0.05	0.10	0.04	0.10	0.03	0.09	0.02	0.10
ITOI	-0.06	0.25	-0.15	0.08	0.07	0.08	0.06	0.08	0.00	0.08
GPA	-0.13	0.27	-0.20*	0.10	0.07	0.09	0.05	0.09	0.17	0.10
$Subject_{CHEM}$	1.93*	0.88	0.10	0.23	-0.28	0.22	-0.25	0.23	0.02	0.24
$Subject_{MATH}$	0.62	1.07	0.04	0.30	0.03	0.29	0.03	0.29	0.00	0.31
$Subject_{PHYS}$	0.99	0.95	-0.14	0.24	-0.24	0.23	-0.21	0.24	0.01	0.25

Female	0.48	0.68	-0.01	0.20	-0.20	0.20	—	—	-0.05	0.21
URM	0.02	0.64	0.06	0.20	0.06	0.20	—	—	0.26	0.21
Year$_{SOPH}$	-0.66	0.84	0.04	0.27	-0.26	0.26	-0.27	0.26	0.39	0.28
Year$_{JUNIOR}$	-0.06	1.03	0.49	0.36	-0.34	0.35	-0.35	0.35	0.37	0.37
Year$_{SENIOR}$	-0.49	0.65	0.24	0.22	0.05	0.22	0.05	0.22	0.12	0.23
At Risk	—		—		—		-0.22	0.25	—	
P-TOI ⨉ At Risk	—		—		—		-0.14	0.24	—	
Academic	—		—		—		—		0.31***	0.09

Note. P-TOI = Perceived Theories of Intelligence; ITOI = Implicit Theories of Intelligence; URM = Underrepresented Racial Minority. All betas are standardized; all SEs are heteroscedasticity corrected. Female and URM not included in model H1d due to dependence with 'at risk' status.
**p* < .05. ***p* < .01. ****p* < .001.

Primary Result After Additional Exclusions

Below we present the results for the tests of primary results after limiting the dataset to students who were enrolled in courses with single instructors (n = 247). Because courses with multiple instructors were primarily biology courses, this resulted in only 8 biology students in this subsample; thus, do not control for subject in this analysis. The results from this analysis are also consistent with the main results.

(a) H1a stated that students who scored higher on Perceived Theories of Intelligence (P-TOI) instrument would be less likely to drop or consider dropping the course during the semester. After controlling for covariates, each unit increase on P-TOI was associated with 41% lower odds of dropping or considering dropping the course, χ^2 (1) = 4.89, n = 133, p = 0.03, OR = 0.59, 95% CI [0.36, 0.94], $r_{partial}$ = -.58.
(b) Our second hypothesis, H1b, was that students who score higher on P-TOI would be less concerned about being evaluated negatively by instructors and peers. We found that for each unit increase on P-TOI, students were, on average, 0.18 units less concerned about being evaluated negatively, χ^2 (1) = 3.77, n = 131, p = .04, $r_{partial}$ = -0.18.
(c) Consistent with the main results, we found that for each unit increase on P-TOI, students reported experiencing, on average, 0.33 units higher course engagement, χ^2 (1) = 13.04, n = 130, p < .001, $r_{partial}$ = 0.35.
(d) We did not find evidence in support of H1d; that is, the interaction between P-TOI and at-risk status was not significant when predicting course engagement, χ^2(1) = 0.04, n = 130, p = 0.82, $r_{partial}$ = 0.02).
(e) Further, scoring higher on P-TOI was not positively correlated with academic help-seeking, χ^2(1) = 0.04, n = 131, p = 0.83, $r_{partial}$ = 0.02. Thus, contrary to our H2 prediction, scores on P-TOI did not predict actual help-seeking behavior during the initial weeks of the course.

As in the primary analyses, the first three hypotheses were empirically supported: students who scored higher on P-TOI were less concerned about negative evaluation from instructors and classmates, more engaged, and less likely to drop or consider dropping the course. The results did not indicate that students at risk for attrition experienced higher course engagement as a function of P-TOI, or that P-TOI was associated with help-seeking behavior during the initial weeks of the semester.

Table D3

Multiple Regression Estimates from Primary Analysis in Study 3 After Additional Exclusions

	H1a: Dropping		H1b: Evaluative Concern		H1c: Course Engagement		H1d: Course Engagement × At Risk		H2: Help-seeking	
	β	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Intercept	-1.64*	0.74	-0.09	0.25	0.40	0.23	0.52	0.27	0.07	0.24
P-TOI (S)	-0.52*	0.24	-0.18*	0.09	0.33***	0.08	0.37	0.21	0.02	0.09
ITOI	-0.18	0.25	-0.14	0.09	0.09	0.09	0.08	0.09	0.00	0.09
GPA	0.20	0.28	-0.20*	0.10	0.08	0.09	0.08	0.09	0.16	0.10
Female	0.61	0.70	-0.06	0.23	-0.35	0.21	—	—	-0.10	0.22
URM	0.31	0.58	0.33	0.22	-0.11	0.21	—	—	0.12	0.22
$\text{Year}_{\text{SOPH}}$	-0.70	0.52	0.26	0.20	-0.02	0.18	-0.01	0.18	-0.09	0.19
$\text{Year}_{\text{JUNIOR}}$	-1.11	0.72	0.05	0.24	-0.36	0.23	-0.33	0.22	0.16	0.24
$\text{Year}_{\text{SENIOR}}$	-0.74	1.18	0.24	0.40	-0.50	0.37	-0.49	0.36	0.70	0.39
At Risk	—		—		—		-0.49	0.27	—	
P-TOI × At Risk	—		—		—		-0.05	0.22	—	
Academic Difficulty	—		—		—		—		0.40***	0.09

Note. P-TOI = Perceived Theories of Intelligence; ITOI = Implicit Theories of Intelligence; URM = Underrepresented Racial Minority. All betas are standardized. Female and URM not included in model H1d due to dependence with At Risk status.

*p < 0.05. ***p < 0.01. **p < 0.01.