ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/287205290

Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high…

Article *in* Journal of Educational Psychology · November 2016

CITATIONS
4

READS
1,969

14 authors, including:

David S. Yeager
University of Texas at Austin
42 PUBLICATIONS **1,548** CITATIONS
SEE PROFILE

Chris Hulleman
University of Virginia
48 PUBLICATIONS **2,015** CITATIONS
SEE PROFILE

Gregory M Walton
Stanford University
38 PUBLICATIONS **2,244** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Online panels data quality View project

Project Growth Mindset Interventions View project

All content following this page was uploaded by Chris Hulleman on 06 January 2016.

The user has requested enhancement of the downloaded file.

Using Design Thinking to Improve Psychological Interventions: The Case of the Growth Mindset During the Transition to High School

David S. Yeager
University of Texas at Austin

Carissa Romero
Dave Paunesku
PERTS and Stanford University

Christopher S. Hulleman
University of Virginia

Barbara Schneider
Michigan State University

Cintia Hinojosa
Hae Yeon Lee
Joseph O'Brien
University of Texas at Austin

Kate Flint
Alice Roberts
Jill Trott
ICF International

Daniel Greene
Gregory M. Walton
Carol S. Dweck
Stanford University

Author Note

The authors would like to thank the teachers, principals, administrators, and students who participated in this research. This research was supported by generous funding from the Spencer Foundation, the William T. Grant Foundation, the Bezos Foundation, the Houston Endowment, the Character Lab, the President and Dean of Humanities and Social Sciences at Stanford University, Angela Duckworth, a William T. Grant scholars award, and a fellowship from the Center for Advanced Study in the Behavioral Sciences (CASBS) to the first and fourth authors. The authors are grateful to Angela Duckworth, Elizabeth Tipton, Michael Weiss, Tim Wilson, Robert Crosnoe, Chandra Muller, Ronald Ferguson, Ellen Konar, Elliott Whitney, Paul Hanselman, Jeff Kosovich, Andrew Sokatch, Katharine Meyer, Patricia Chen, Chris Macrander, Jacquie Beaubien, and Rachel Herter for their assistance. The look and feel of the revised intervention was developed by Matthew Kandler. Please address correspondence to David Yeager (dyeager@utexas.edu) or Gregory Walton (gwalton@stanford.edu).

Abstract

There are many promising psychological interventions on the horizon, but there is no clear methodology for preparing them to be scaled up. Drawing on design thinking, the present research formalizes a methodology for redesigning and tailoring initial interventions. We test the methodology using the case of fixed versus growth mindsets during the transition to high school. Qualitative inquiry and rapid, iterative, randomized "A/B" experiments were conducted with ~3,000 participants to inform intervention revisions for this population. Next, two experimental evaluations showed that the revised growth mindset intervention was an improvement over previous versions in terms of short-term proxy outcomes (Study 1, *N*=7,501), and it improved 9th grade core-course GPA and reduced D/F GPAs for lower achieving students when delivered via the Internet under routine conditions with ~95% of students at 10 schools (Study 2, *N*=3,676). Although the intervention could still be improved even further, the current research provides a model for how to improve and scale interventions that begin to address pressing educational problems. It also provides insight into how to teach a growth mindset more effectively.

Keywords: motivation, psychological intervention, incremental theory of intelligence, growth mindset, adolescence.

Using Design Thinking to Improve Psychological Interventions:
The Case of the Growth Mindset During the Transition to High School

One of the most promising developments in educational psychology in recent years has been the finding that self-administered psychological interventions can initiate lasting improvements in student achievement (Cohen & Sherman, 2014; Garcia & Cohen, 2012; Walton, 2014; Wilson, 2012; Yeager & Walton, 2011). These interventions do not provide new instructional materials or pedagogies. Instead, they capitalize on the insights of expert teachers (see Lepper & Woolverton, 2001; Treisman, 1992) by addressing students' subjective *construals* of themselves and school—how students view their abilities, their experiences in school, their relationships with peers and teachers, and their learning tasks (see Ross & Nisbett, 1991).

For instance, students can show greater motivation to learn when they are led to construe their learning situation as one in which they have the potential to develop their abilities (Dweck, 1999; Dweck, 2006), in which they feel psychologically safe and connected to others (Cohen et al., 2006; Stephens, Hamedani, & Destin, 2014; Walton & Cohen, 2007), and in which putting forth effort has meaning and value (Hulleman & Harackiewicz, 2009; Yeager et al., 2014; see Eccles & Wigfield, 2002; also see Elliot & Dweck, 2005; Lepper, Woolverton, Mumme, & Gurtner, 1993; Stipek, 2002). Such subjective construals—and interventions or teacher practices that affect them—can affect behavior over time because they can become self-confirming. When students doubt their capacities in school—for example, when they see a failed math test as evidence that they are not a "math person"—they behave in ways that can make this true, for example, by studying less rather than more or by avoiding future math challenges they might learn from. By changing initial construals and behaviors, psychological interventions can set in motion recursive processes that alter students' achievement into the future (see Cohen & Sherman, 2014; Garcia & Cohen, 2012; Walton, 2014; Yeager & Walton, 2011).

Although promising, self-administered psychological interventions have not often been tested in ways that are sufficiently relevant for policy and practice. For example, rigorous randomized trials have been conducted with only limited samples of students within schools—those who could be conveniently recruited. These studies have been extremely useful for testing of novel theoretical claims (e.g., Aronson et al., 2002; Blackwell et al., 2007; Good et al., 2003). Some studies have subsequently taken a step toward scale by developing methods for delivering intervention materials to large samples via the Internet without requiring professional development (e.g. Paunesku et al., 2015; Yeager et al., 2014). However, such tests are limited in relevance for policy and practice because they did not attempt to improve the outcomes of an entire student body or entire sub-groups of students.

There is not currently a methodology for adapting materials that were effective in initial experiments so they can be improved and made more effective for populations of students who are facing particular issues at specific points in their academic lives. We seek to develop this methodology here. To do so, we focus on students at diverse schools but at a similar developmental stage, and who therefore may encounter similar academic and psychological challenges and may benefit from an intervention to help them navigate those challenges. We test whether the tradition of "design thinking," combined with advances in psychological theory, can facilitate the development of improved intervention materials for a given population.[1]

As explained later, the policy problem we address is *core academic performance of 9th graders transitioning to public high schools in the United States* and the specific intervention we re-design is the *growth mindset of intelligence* intervention (also called an incremental theory of intelligence intervention; Aronson et al., 2002; Blackwell et al., 2007; Good et al., 2003;

[1] Previous research has shown how design thinking can improve instructional materials but not yet psychological interventions (see Razzouk & Shute, 2012).

Paunesku et al., 2015). The growth mindset intervention counteracts the *fixed mindset* (also called an entity theory of intelligence), which is the belief that intelligence is a fixed entity that cannot be changed with experience and learning. The intervention teaches scientific facts about the malleability of the brain, to show how intelligence can be developed. It then uses writing assignments to help students internalize the messages (see the pilot study methods for detail). The growth mindset intervention aims to increase students' desires to take on challenges and to enhance their persistence, by forestalling attributions that academic struggles and setbacks mean one is "not smart" (Blackwell et al., 2007, Study 1; see Burnette et al., 2013; Yeager & Dweck, 2012). These psychological processes can result in academic resilience.

Design Thinking and Psychological Interventions

To develop psychological interventions, expertise in theory is crucial. But theory alone does not help a designer discover how to connect with students facing a particular set of motivational barriers. Doing that, we believe, is easier when combining theoretical expertise with a design-based approach (Razzouk & Shute, 2012; also see Bryk, 2009; Yeager & Walton, 2011). Design thinking is "problem-centered." That is, effective design seeks to solve predictable problems for specified user groups (Kelley & Kelley, 2013; also see Bryk, 2009; Razzouk & Shute, 2012). Our hypothesis is that this problem-specific customization, guided by theory, can increase the likelihood that an intervention will be more effective for a pre-defined population.

We apply two design traditions, *user-centered design* and *A/B testing*. Theories of user-centered design were pioneered by firms such as IDEO (Kelley & Kelley, 2013; see Razzouk & Shute, 2012). The design process privileges the user's subjective perspective—in the present case, 9th grade students. To do so, it often employs qualitative research methods such as ethnographic observations of people's mundane goal pursuit in their natural habitats (Kelley &

Kelley, 2013). User-centered design also has a bias toward action. Designers test minimally viable products early in the design phase in an effort to learn from users how to improve them (see Ries, 2011). Applied to psychological intervention, this can help prevent running a large, costly experiment with an intervention that has easily discoverable flaws. In sum, our aim was to acquire insights about the barriers to students' adoption of a growth mindset during the transition to high school as quickly as possible, using data that were readily obtainable, without waiting for a full-scale, long-term evaluation.

User-centered design typically does *not* ask users what they desire or would find compelling. Individuals may not always have access to that information (Wilson, 2002). However, users may be excellent reporters on what they dislike or are confused by. Thus, user-centered designers do not often ask, "What kind of software would you like?" Rather, they often show users prototypes and let them say what seems wrong or right. Similarly, we did not ask students, "What would make you adopt a growth mindset?" But we did ask for positive and negative reactions to prototypes of growth mindset materials. We then used those responses to formulate changes for the next iteration.

Qualitative user-centered design can lead to novel insights, but how would one know if those insights were actually improvements? Experimentation can help. Therefore we drew on a second tradition, that of "*A/B testing*" (see, e.g., Kohavi & Longbotham, 2015). The logic is simple. Because it is easy to be wrong about what will be persuasive to a user, rather than guess, test. We used the methodology of low-cost, short-term, large-sample, random-assignment experiments to test revisions to intervention content. Although each experiment may not, on its own, offer a theoretical advance or a definitive conclusion, in the aggregate they may improve an intervention for a population of interest.

Showing that this design process has produced intervention materials that are more ready for scaling requires meeting at least two conditions: (1) the re-designed intervention should be more effective for the target population than a previous iteration when examining short-term proxy outcomes, and (2) the re-designed intervention should address the policy-relevant aim: namely, it should benefit student achievement when delivered to entire populations of students within schools. Interestingly, although a great deal has been written about best practices in design (e.g. Razzouk & Shute, 2012), we do not know of any set of experiments that has evaluated the end-product of a design process by collecting both types of evidence.

Focus of the Present Investigation

As a first test case, we redesign a growth mindset of intelligence intervention to improve it and to make it more effective for the transition to high school (Aronson et al., 2002; Good et al., 2003; Blackwell et al., 2007; Paunesku et al., 2015; see Dweck, 2006; Yeager & Dweck, 2012). This is an informative case study because (a) previous research has found that growth mindsets can predict success across educational transitions, and previous growth mindset interventions have shown some effectiveness; (b) there is a clearly defined psychological process model explaining how a growth mindset relates to student performance supported by a large amount of correlational and laboratory experimental research (see a meta-analysis by Burnette et al., 2013), which informs decisions about which "proxy" measures would be most useful for shorter-term evaluations. The growth mindset is thus a good place to start.

As noted, our defined user group was students making the transition to high school. New 9th graders represent a large population of students—approximately 4 million individuals each year (Bauman & Davis, 2013). Although entering 9th graders' experiences can vary, there are some common challenges: high school freshmen often take more rigorous classes than

previously and their performance can affect their chances for college; they have to form relationships with new teachers and school staff; and they have to think more seriously about their goals in life. Students who do not successfully complete 9th grade core courses have a dramatically lower rate of high school graduation, and much poorer life prospects (Allensworth & Easton, 2005). Improving the transition to high school is an important policy objective.

Previous growth mindset interventions might be better tailored for the transition to high school in several ways. First, past growth mindset interventions were not designed for the specific challenges that occur in the transition to high school. Rather they were often written to address challenges that a learner in any context might face, or they were aimed at middle school (Blackwell et al., 2007; Good et al., 2003) or college (Aronson et al., 2002). Second, interventions were not written for the vocabulary, conceptual sophistication, and interests of adolescents entering high school. Third, when they were created, they did not have in mind arguments that might be most relevant or persuasive for 14-15 year-olds.

Despite this potential lack of fit, prior growth mindset interventions have already shown to have initial effectiveness in high schools. Paunesku et al. (2015) conducted a double-blind, randomized experimental evaluation of a growth mindset intervention with over 1,500 high school students via the Internet. The authors found a significant Intervention × Prior achievement interaction, such that lower-performing students benefitted most from the growth mindset intervention in terms of their GPA. Lower-achievers both may have had more room to grow (given range restriction for higher achievers) and also may have faced greater concerns about their academic ability (see analogous benefits for lower-achievers in Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009; Hulleman & Haraciewicz, 2009; Wilson & Linville, 1982; Yeager, Henderson, et al., 2014). Looking at grade improvement with a different measure,

Paunesku et al. also found that previously low-achieving treated students were also less likely to receive "D" and "F" grades in core classes (e.g., English, math, science). We examine whether these results could be replicated with a re-designed intervention.

Overview of Studies

The present research, first, involved design thinking to create a newer version of a growth mindset intervention, presented here as a pilot study. Next, Study 1 tested whether the design process produced growth mindset materials that were an improvement over original materials when examining proxy outcomes, such as beliefs, goals, attributions, and challenge-seeking behavior (see Blackwell et al., 2007 or Burnette et al., 2013 for justification). The study required a great deal of power to detect a significant intervention contrast because the generic intervention (called the "original" intervention here; Paunesku et al., 2015) also taught a growth mindset.

Note that we did not view the re-designed intervention as "final," as ready for universal scaling, or as representing the best possible version of a growth mindset. Instead our goal was more modest: we simply sought to document that a design process could create materials that were a significant improvement relative to their predecessor.

Study 2 tested whether we had developed a revised growth mindset intervention that was actually effective at changing achievement when delivered to a census of students (>95%) in 10 different schools across the country.[2] The focal research hypothesis was an effect on core course GPA and D/F averages among previously lower-performing students, which would replicate the Intervention × Prior achievement interaction found by Paunesku et al. (2015).

Study 2 was, to our knowledge, the first pre-registered replication of a psychological intervention effect, the first to have data collected and cleaned by an independent research firm,

[2] A census is defined as an attempt to reach all individuals in an organization, and is contrasted with a sample, which does not.

and the first to employ a census (>95% response rates). Hence, it was a rigorous test of the hypothesis.

Pilot: Using Design Thinking to Improve a Growth Mindset Intervention

Method

Data. During the design phase, the goal was to learn as much as possible, as rapidly as possible, given data that were readily available. Thus, no intentional sampling was done and no demographic data on participants were collected. For informal qualitative data—focus groups, one-on-one interviews, and other types of feedback—high school students were contacted through personal connections and invited to provide feedback on a new program for high school students.

Quantitative data—the rapid "A/B" experiments—were collected from college-aged and older adults on Amazon's Mechanical Turk platform. Although adults are not an ideal data source for a transition to high school intervention, the A/B experiments required great statistical power because we were testing minor variations. Data from Mechanical Turk provided this and allowed us to iterate quickly. Conclusions from those tests were tempered by the qualitative feedback from high school students. At the end of this process, we conducted a final rapid A/B randomized experiment with high school students using the Qualtrics user panel.

Materials.

The "original" mindset intervention. The original intervention was the starting point for our revision. It involved three elements. First, participants read a scientific article titled "You Can Grow Your Intelligence," written by researchers, used in the Blackwell et al. (2007) experiment, and slightly revised for the Paunesku et al. (2015) experiments. It described the idea that the brain can get smarter the more it is challenged, like a muscle. As scientific background

for this idea, the article explained what neurons are and how they form a network in the brain. It then provided summaries of studies showing that animals or people (e.g., rats, babies, or London taxi drivers) who have greater experience or learning develop denser networks of neurons in their brains.

After reading this four-page article, participants were asked to generate a personal example of learning and getting smarter—a time when they used to not know something, but then they practiced and got better at it. Finally, participants were asked to author a letter encouraging a future student who might be struggling in school and may feel "dumb." This is a "saying-is-believing" exercise (E. Aronson, 1999; J. Aronson et al., 2002; Walton & Cohen, 2011; Walton, 2014).

"Saying-is-believing" is thought to be effective for several reasons. First, it is thought to make the information (in this case, about the brain and its ability to grow) more self-relevant, which may make it easier to recall (Bower & Gilligan, 1979; Hulleman & Harackiewicz, 2009; Lord, 1980). Prior research has found that students can benefit more from social-psychological intervention materials when they author reasons why the content is relevant, as opposed to being told why they are relevant to their own lives (Godes, Hulleman, & Harackiewicz, 2007). Second, by mentally rehearsing how one should respond when struggling, it can be easier to enact those thoughts or behaviors later (Gollwitzer, 1999). Third, when students are asked to communicate the message to someone else—and not directly asked to believe it themselves—it can feel less controlling, it can avoid implying that students are deficient, and it can lead students to convince themselves of the truth of the proposition via cognitive dissonance processes (see research on self-persuasion, Aronson, 1999; also see Bem, 1965; Cooper & Fazio, 1984).

Procedures and Results: Revising the Mindset Intervention

Design methodology.

User-centered design. We (a) met one-on-one with 9th graders; (b) met with groups of 2 to 10 students; and (c) piloted with groups of 20-25 9th graders. In these piloting sessions we first asked students to go through a "minimally viable" version of the intervention (i.e. early draft revisions of the "original" materials) as if they were receiving it as a new freshman in high school. We then led them in guided discussions of what they disliked, what they liked, and what was confusing. We also asked students to summarize the content of the message back to us, under the assumption that a person's inaccurate summary of a message is an indication of where the clarity could be improved. The informal feedback was suggestive, not definitive. However a consistent message from the students allowed the research team to identify, in advance, predictable failures of the message to connect or instruct, and potential improvements worth testing.

Through this we developed insights that might seem minor taken individually, but, in the aggregate, may be important. These were: to include quotes from admired adults and celebrities; to include more and more diverse writing exercises; to weave purposes for *why* one should grow one's brain together with statements that one *could* grow one's brain; to use bullet points instead of paragraphs; to reduce the amount of information on each page; to show actual data from past scientific research in figures rather than summarize them generically (because it felt more respectful); to change examples that appear less relevant to high school students (e.g., replacing a study about rats growing their brains with a summary of science about teenagers' brains), and more.

A/B testing. A series of randomized experiments tested specific variations on the mindset intervention. In each, we randomized participants to versions of a mindset intervention and

assessed changes from pre- to post-test in self-reported fixed mindsets (see measures below in Study 1; also see Dweck, 1999). Mindset self-reports are an imperfect measure, as will be shown in the two studies. Yet they are informative in the sense that if mindsets are *not* changed—or if they were changed in the wrong direction—then it is reason for concern. Self-reports give the data "a chance to object" (cf. Latour, 2005)

Two studies involved a total of 7 factors, fully crossed, testing 5 research questions across *N*=3,004 participants. These factors and their effects on self-reported fixed mindsets are summarized in Table 1.

One question tested in both A/B experiments was whether it was more effective to tell research participants that the growth mindset intervention was designed to help them (a "direct" framing), versus a framing in which participants were asked to help evaluate and contribute content for future 9th grade students (an "indirect" framing). Research on the "saying-is-believing" tactic (E. Aronson, 1999; J. Aronson et al., 2002; Walton & Cohen, 2011; Yeager & Walton, 2011) and theory about "stealth" interventions more broadly (Robinson, 2010) suggest that the latter might be more helpful, as noted above. Indeed, Table 1 shows that in both A/B Study 1 and A/B Study 2 the direct framing led to *smaller* changes in mindsets—corresponding to lower effectiveness—than indirect framing (see rows 1 and 5 in Table 1). Thus, the indirect framing was used throughout the revised intervention. To our knowledge this is the first experimental test of the effectiveness of the indirect framing in psychological interventions, even though it is often standard practice (J. Aronson et al., 2002; Walton & Cohen, 2011; Yeager & Walton, 2011).

The second question was: Is it more effective to present and refute the fixed mindset view? Or is it more effective to only teach evidence for the growth mindset view? On the one

hand, refuting the fixed mindset might more directly discredit the problematic belief. On the other hand, it might give credence and voice to the fixed mindset message, for instance by conveying that the fixed mindset is a reasonable perspective to hold (perhaps even the norm), giving it an "illusion of truth" (Skurnik, Yoon, Park, & Schwarz, 2005). This might cause participants who hold a fixed mindset to become entrenched in their beliefs. Consistent with the latter possibility, in A/B Study 1, refuting the fixed mindset view led to smaller changes in mindsets—corresponding to lower effectiveness—as compared to not doing so (see row 2 in Table 1). Furthermore, the refutation manipulation caused participants who held a stronger fixed mindset at baseline to show an *increase* in fixed mindset post-message, main effect $p = .003$, interaction effect $p = .01$. That is, refuting a fixed mindset seemed to exacerbate fixed mindset beliefs for those who already held them.

Following this discovery, we wrote a more subtle and indirect refutation of fixed mindset thinking and tested it in A/B Study 2. The revised content encouraged participants to replace thoughts about *between-person* comparisons (that person is smarter than me) with *within-person* comparisons (I can become even smarter tomorrow than I am today). This no longer caused reduced effectiveness (see row 6 in Table 1). The final version emphasized within-person comparisons as a means to discrediting between-person comparisons. As an additional precaution, beyond what we had tested, the final materials never named or defined a "fixed mindset."

We additionally tested the impact of using well-known or successful adults as role models of a growth mindset. For instance, the intervention conveyed the true story of Scott Forstall, who, with his team, developed the first iPhone at Apple. Forstall used growth mindset research to select team members who were not afraid of failure but were ready for a challenge. It

furthermore included an audio excerpt from a speech given by First Lady Michelle Obama, in which she summarized the basic concepts of growth mindset research. These increased adoption of a growth mindset compared to versions that did not include these endorsements (Table 1). Other elements were tested as well (see Table 1).

Guided by theory. The re-design also relied on psychological theory. These changes, which were not subjected to A/B testing, are summarized here. Several theoretical elements were relevant: (a) theories about how growth mindset beliefs affect student behavior in practice; (b) theories of cultural values that may appear to be in contrast with growth mindset messages; and (c) theories of how best to induce internalization of attitudes among adolescents.

Emphasizing "strategies," not just "hard work." We were concerned that the "original" intervention too greatly emphasized "hard work" as the opposite of raw ability, and under-emphasized the need to change strategies or ask adults for advice on improved strategies for learning. This is because just working harder with ineffective strategies will not lead to increased learning. So, for example, the revised intervention said "Sometimes people want to learn something challenging, and they try hard. But they get stuck. That's when they need to try new strategies—new ways to approach the problem" (also see Yeager & Dweck, 2012). Our goal was to remove any stigma of needing to ask for help or having to switch one's approach.

Addressing a culture of independence. We were concerned that the notion that *you* can grow *your* intelligence would perhaps be perceived as too "independent," and threaten the more communal, interdependent values that many students might emphasize, especially students from working class backgrounds and some racial/ethnic minority groups (Fryberg, Covarrubias, & Burack, 2013; Stephens et al., 2014). Therefore we included more prosocial, beyond-the-self

motives for adopting and using a growth mindset (see Hulleman & Harackiewicz, 2009; Yeager et al., 2014). For example, the new intervention said:

> "People tell us that they are excited to learn about a growth mindset because it helps them achieve the goals that matter to them and to people they care about. They use the mindset to learn in school so they can give back to the community and make a difference in the world later."

Aligning norms. Adolescents may be especially likely to conform to peers (Cohen & Prinstein, 2006), and so we created a norm around the use of a growth mindset (Cialdini et al., 1991). For instance, the end of the second session said "People everywhere are working to become smarter. They are starting to understand that struggling and learning are what put them on a path to where they want to go." The intervention furthermore presented a series of stories from older peers who endorsed the growth mindset concepts.

Harnessing reactance. We sought to use adolescent *reactance*, or the tendency to reject mainstream or external exhortations to change personal choices (Brehm, 1966; Erikson, 1968; Hasebe, Nucci, & Nucci, 2004; Nucci, Killen, & Smetana, 1996), as an asset rather than as a source of resistance to the message. We did this by initially framing the mindset message as a reaction to adult control. For instance, at the very beginning of the intervention adolescents read this story from an upper year student:

> "I hate how people put you in a box and say 'you're smart at this' or 'not smart at that." After this program, I realized the truth about labels: they're made up. … now I don't let other people box me in … it's up to me to put in the work to strengthen my brain."

Self-persuasion. The revised intervention increased the number of opportunities for participants to write their own opinions and stories. This was believed to increase the probability that more of the benefits of "saying-is-believing" would be achieved.

Study 1: Does a Revised Growth Mindset Intervention Outperform an Existing Effective Intervention?

Study 1 evaluated whether the design process resulted in materials that were an improvement over the originals. As criteria, Study 1 used short-term measures of psychological processes that are well-established to follow from a growth mindset: person versus process-focused attributions for difficulty (low ability vs. strategy or effort), performance avoidance goals (over-concern about making mistakes or looking incompetent), and challenge-seeking behavior (Blackwell et al., 2007; Mueller & Dweck, 1998; see Burnette et al., 2013; Dweck & Leggett, 1988; Yeager & Dweck, 2012). Note that the goal of this research was not to test whether any individual revision, by itself, caused greater efficacy, but rather to investigate all of the changes, in the aggregate, in comparison to the original intervention.

Method

Data. A total of 69 high schools in the United States and Canada were recruited, and 7,501 9th grade students (predominately ages 14-15) provided data during the session when dependent measures were collected (Time 2), although not all finished the session.[3] Participants were diverse: 17% were Hispanic/Latino, 6% were black/African-American, 3% were Native American / American Indian, 48% were White, non-Hispanic, 5% were Asian / Asian-American,

[3] This experiment involved a third condition of equal cell size that was developed to test other questions. Results involving that condition will be presented in a future report, but they are fully consistent with all of the findings presented here.

and the rest were from another or multiple racial groups. Forty-eight percent were female, and 53% reported that their mothers had earned a Bachelor's degree or greater.

Procedures.

School recruitment. Schools were recruited via advertisements in prominent educational publications, through social media (e.g. Twitter), and through recruitment talks to school districts or other school administrators. Schools were informed that they would have access to a free growth mindset intervention for their students. Because of this recruitment strategy, *all* participating students indeed received a version of a growth mindset intervention. The focal comparison was between an "original" version of a growth mindset intervention (Paunesku et al., 2015, which was adapted from materials in Blackwell et al., 2007), and a "revised" version, created via the user-centered, rapid-prototyping design process summarized above. See screenshots in Figure 1.

The original intervention. Minor revisions were made to the original intervention to make it more parallel to the revised intervention, such as the option for the text to be read to students.

Survey sessions. In the winter of 2015, school coordinators brought their students to the computer labs for two sessions, 1 to 4 weeks apart (researchers never visited the schools). The Time 1 session involved baseline survey items, a randomized mindset intervention, some fidelity measures, and brief demographics. Random assignment happened in real time and was conducted by a web server, and so all staff were blind to condition. The Time 2 session involved a second round of content for the revised mindset intervention, and control exercises for the original mindset condition. At the end of session two, students completed proxy outcome measures.

Measures. In order to minimize respondent burden and increase efficiency at scale, we used or developed 1- to 3-item self-report measures of our focal constructs (see a discussion of "practical measurement" in Yeager & Bryk, 2015). We also developed a brief behavioral task.

Fixed mindset. Three items at Time 1 and Time 2 assessed fixed mindsets: "You have a certain amount of intelligence, and you really can't do much to change it," "Your intelligence is something about you that you can't change very much," and "Being a "math person" or not is something that you really can't change. Some people are good at math and other people aren't." (Response options: 1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Mostly disagree*, 4 = *Mostly agree*, 5 = *Agree*, 6 = *Strongly agree*). These were averaged into a single scale with higher values corresponding to more fixed mindsets (α=.74).

Challenge-seeking: the "Make-a-Math-Worksheet" Task. We created a novel behavioral task to assess a known behavioral consequence of a growth mindset: challenge-seeking (Blackwell et al., 2007; Mueller & Dweck, 1998). This task may allow for a detection of effects for previously high-achieving students, because GPA may have a range restriction at the high end. It used Algebra and Geometry problems obtained from the Khan Academy website and was designed to have more options than previous measures (Mueller & Dweck, 1998) to produce a continuous measure of challenge seeking. Participants read these instructions:

> "We are interested in what kinds of problems high school math students prefer to work on. On the next few pages, we would like you to create your own math worksheet. If there is time, at the end of the survey you will have the opportunity to answer these math problems. On the next few pages there are problems from 4 different math chapters. Choose between 2 and 6 problems for each chapter.

> You can choose from problems that are...**Very challenging** but you might learn a lot; **Somewhat challenging** and you might learn a medium amount; **Not very challenging** and you probably won't learn very much; Do not try to answer the math problems. Just click on the problems you'd like to try later if there's time."

See Figure 2. There were three topic areas (Introduction to Algebra, Advanced Algebra, and Geometry), and within each topic area there were four "chapters" (e.g. rational and irrational numbers, quadratic equations, etc.), and within each chapter there were six problems, each labeled "Not very challenging," "Somewhat challenging," or "Very challenging" (two per type).[4] Each page showed the six problems for a given chapter and, as noted, students were instructed to select "at least 2 and up to 6 problems" on each page.

The total number of "Very challenging" (i.e. *hard*) problems chosen across the 12 pages was calculated for each student (Range: 0-24) as was the total number of "Not very challenging" (i.e. *easy*) problems (Range: 0-24). The final measure was the number of easy problems minus the number of hard problems selected. Visually, the final measure approximated a normal distribution.

Challenge-seeking: Hypothetical scenario. Participants were presented with the following scenario, based on a measure in Mueller and Dweck (1998):

> "Imagine that, later today or tomorrow, your math teacher hands out two extra credit assignments. You get to choose which one to do. You get the same number of points for trying either one. One choice is an easy review—it has math problems you already know how to solve, and you will probably get most of the answers right without having to think

[4] A screening question also asked students to list what math course they are currently taking. Supplementary analyses found results were no different when limiting the worksheet task data only to the "chapters" that correspond to the course a student was enrolled in.

very much. It takes 30 minutes. The other choice is a hard challenge—it has math problems you don't know how to solve, and you will probably get most of the problems wrong, but you might learn something new. It also takes 30 minutes. If you had to pick right now, which would you pick?"

Participants chose one of two options (1 = *The easy math assignment where I would get most problems right*, 0 = *The hard math assignment where I would possibly learn something new*). Higher values corresponded to the avoidance of challenge, and so this measure should be positively correlated with fixed mindset and be reduced by the mindset intervention.

Fixed-trait attributions. Fixed mindset beliefs are known predictors of person-focused versus process-focused attributional styles (e.g., Henderson & Dweck, 1990; Robins & Pals, 2002). We adapted prior measures (Blackwell et al., 2007) to develop a briefer assessment. Participants read this scenario: "Pretend that, later today or tomorrow, you got a bad grade on a very important math assignment. Honestly, if that happened, how likely would you be to think these thoughts?" Participants then rated this fixed-trait, person-focused response "This means I'm probably not very smart at math" and this malleable, process-focused response "I can get a higher score next time if I find a better way to study (reverse-scored)" (response options: 1 = *Not at all likely*, 2 = *Slightly likely*, 3 = *Somewhat likely*, 4 = *Very likely*, 5 = *Extremely likely*). The two items were averaged into a single composite, with higher values corresponding to more fixed-trait, person-focused attributional responses.

Performance avoidance goals. Because fixed mindset beliefs are known to predict the goal of hiding one's lack of knowledge (Dweck & Leggett, 1988), we measured performance-avoidance goals with a single item (Elliot & McGregor, 2001; performance approach goals were not measured). Participants read "What are your goals in school from now until the end of the

year? Below, say how much you agree or disagree with this statement. One of my main goals for the rest of the school year is to avoid looking stupid in my classes" (Response options: 1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Mostly disagree*, 4 = *Mostly agree*, 5 = *Agree*, 6 = *Strongly agree*). Higher values correspond to greater performance avoidance goals.

Fidelity measures. To examine fidelity of implementation across conditions, students were asked to report on distraction in the classroom, both *peers*' distraction ("Consider the students around you... How many students would you say were working carefully and quietly on this activity today?" Response options: 1= *Fewer than half of students*, 2 = *About half of students*, 3 = *Most students*, 4 = *Almost all students, with just a few exceptions*, 5 = *All students*) and one's *own distraction* ("How distracted were you, personally, by other students in the room as you completed this activity today?" Response options: 1 = *Not distracted at all*, 2 = *Slightly distracte*d, 3 = *Somewhat distracted*, 4 = *Very distracted*, 5 = *Extremely distracted*).

Next, participants in both conditions rated how *interesting* the materials were ("For you personally, how interesting was the activity you completed in this period today?" Response options: 1 = *Not interesting at all*, 2 = *Slightly interesting*, 3 = *Somewhat interesting*, 4 = *Very interesting*, 5 = *Extremely interesting*), and how much they *learned* from the materials ("How much do you feel that you learned from the activity you completed in this period today?" Response options: 1 = *Nothing at all*, 2 = *A little*, 3 = *A medium amount*, 4 = *A lot*, 5 = *An extreme amount*).

Prior achievement. School records were not available for this sample. Prior achievement was indexed by a composite of self-reports of typical grades and expected grades. The two items were "Thinking about this school year and the last school year, what grades do you usually get in core classes? By core classes, we mean: English, math, and science. We don't mean electives,

like P.E. or art" (Response options: 1 = *Mostly F's*, 2 = *Mostly D's*, 3 = *Mostly C's*, 4 = *Mostly B's*, 5 = *Mostly A's*) and "Thinking about your skills and the difficulty of your classes, how do you think you'll do in math in high school? (Response options: 1 = *Extremely poorly*, 2 = *Very poorly*, 3 = *Somewhat poorly*, 4 = *Neither well nor poorly*, 5 = *Somewhat well*, 6 = *Very well*, 7 = *Extremely well*). Items were z-scored within schools and then averaged, with higher values corresponding to higher prior achievement (α=.74).

Attitudes to validate the "Make-a-Math-Worksheet" Task. Additional self-reports were assessed at Time 1 to validate the Time 2 challenge-seeking behaviors. These were all expected to be correlated with challenge-seeking behavior. These were the short *grit* scale (Duckworth & Quinn, 2009), the academic *self-control* scale (Tsukayama, Duckworth, & Kim, 2013), a single item of *interest* in math ("In your opinion, how interesting is the subject of math in high school?" response options: 1 = *Not at all interesting*, 2 = *Slightly interesting*, 3 = *Somewhat interesting*, 4 = *Very interesting*, 5 = *Extremely interesting*), and a single item of *math anxiety* ("In general, how much does the subject of math in high school make you feel nervous, worried, or full of anxiety?" Response options: 1 = *Not at all*, 2 = *A little*, 3 = *A medium amount*, 4 = *A lot*, 5 = *An extreme amount*).

Results

Preliminary analyses. We tested for violations of assumptions of linear models (e.g., outliers, non-linearity). Variables either did not violate linearity, or when transforming or dropping outliers, significance of results were unchanged.

Correlational analyses. Before examining the impact of the intervention, we conducted correlational tests to replicate the basic findings from prior research on fixed versus growth mindsets. Table 2 shows that measured fixed mindset significantly predicted fixed-trait, person-

focused attributions, $r(6636)=.28$, performance avoidance goals, $r(6636)=.23$, and hypothetically choosing easy problems over hard problems, $r(6636)=.12$ (all ps < .001). The sizes of these correlations correspond to the sizes in a meta-analysis of many past studies (Burnette et al., 2013). Thus, a fixed mindset was associated with thinking that difficulty means you are "not smart," with having the goal of not looking "dumb," and with avoiding hard problems that you might get wrong, as in prior research (see Dweck, 2006; Dweck & Leggett, 1988; Yeager & Dweck, 2012).

Validating the "Make-a-math-worksheet" challenge-seeking task. As shown in Table 2, the choice of a greater number of easy problems as compared to hard problems at Time 2 was modestly correlated with grit, self-control, prior performance, interest in math, math anxiety measured at Time 1, 1 to 4 weeks earlier, and all in the expected direction (all ps<.001, given large sample size; see Table 2).

In addition, measured fixed mindset, fixed-trait attributions, and performance avoidance goals predicted choices of more easy problems and fewer hard problems. The worksheet task behavior correlated with the single dichotomous hypothetical choice. Thus, participants' choices on this task appear to reflect their individual differences in challenge-seeking tendencies.

Random assignment. Random assignment to condition was effective. There were no differences between conditions in terms of demographics (gender, race, ethnicity, special education, parental education) or in terms of prior achievement (all ps>.1), despite 80% power to detect effects as small as d=.06. Furthermore, as shown in Table 3, there were no pre-intervention differences between conditions in terms of fixed mindset.

Fidelity. Students in the revised and original intervention conditions did not differ in terms of their ratings of their peers' distraction during the testing session, $t(6454)=0.35$, p=.72, or

in terms of their own personal distraction, $t(6454)=1.92$, $p=.06$. Although there was a trend toward greater distraction in the original intervention group, this was likely a result of very large sample size. Regardless, distraction was low for both groups (at or below a 2 on a 5-point scale).

Next, the revised intervention was rated as more interesting, $t(6454)=4.44$, $p<.001$, and also as more likely to cause participants to feel as though they learned something, $t(6454)=6.25$, $p<.001$. This means the design process was successful in making the new intervention engaging. Yet this meant that in Study 2 it was important to ensure that the control condition was also interesting.

Self-reported fixed mindset. The revised mindset intervention was more effective at reducing reports of a fixed mindset as compared to the original mindset intervention. See Table 3. The original mindset group showed a change score of Δ =-0.22 scale points (out of 6), as compared to Δ =-0.48 scale point (out of 6) for the revised mindset group. Both change scores were significant at $p<.001$, and they were significantly different from one another (see Table 2).

In moderation analyses, students who already had more of a growth mindset at baseline changed their beliefs less, Intervention × Pre-intervention fixed mindset interaction, $t(6687)=-3.385$, $p=.0007$, $\beta=.04$, which is consistent with a ceiling effect among those who already held a growth mindset. There was no Intervention × Prior achievement interaction, $t(6687)=1.184$, $p=.24$, $\beta=.01$, suggesting that the intervention was effective in changing mindsets across all levels of achievement.

Primary analyses.

The "make-a-math-worksheet" task. Our primary outcome of interest was challenge-seeking behavior. Compared to the original growth mindset intervention, the revised growth mindset intervention reduced the tendency to choose more easy than hard math problems, 4.62

more easy than hard vs. 2.42 more easy than hard, a significant difference, $t(6884)=8.03$, $p<.001$, $d=.19$. See Table 3. This intervention effect on behavior was not moderated by prior achievement, $t(6884)=-.65$, $p=.52$, $\beta=.01$, or pre-intervention fixed mindset, $t(6884)=.63$, $p=.53$, $\beta=.01$, showing that the intervention led high and low achievers alike to demonstrate greater challenge-seeking.

It was also possible to test whether the revised growth mindset intervention increased the overall number of challenging problems, decreased the number of easy problems, or increased the proportion of problems chosen that were challenging. Supplementary analyses showed that all of these intervention contrasts were also significant, $t(6884)=3.95$, $p<.001$, $t(6884)=8.60$, $p<.001$, $t(6884)=7.60$, $p<.001$, respectively.

Secondary analyses.

Hypothetical challenge-seeking scenario. Compared to the original mindset intervention, the revised mindset intervention reduced the proportion of students saying they would choose the "easy" math homework assignment versus the "hard" assignment from 60% to 51%, logistic regression $Z=7.951$, $p<.001$, $d=.19$. See Table 3. The intervention effect was not significantly moderated by prior achievement, $Z=-1.82$, $p=.07$, $\beta=.04$, or pre-intervention fixed mindset, $Z=0.51$, $p=.61$, $\beta=.01$.

Attributions and goals. The revised intervention significantly reduced fixed-trait, person-focused attributions as well as performance avoidance goals, compared to the original intervention, ps<.01 (see Table 3). These effects were small, ds = .07 and 06, but recall that this is the group difference between two growth mindset interventions. Neither of these outcomes showed a significant Intervention × Pre-intervention fixed mindset interaction, $t(6647)=-.62$, $p=.54$, $\beta=.01$, and $t(6631)=-.72$, $p=.47$, $\beta=.01$, for attributions and goals respectively. For

attributions, there was no Intervention × Prior achievement interaction, $t(6647)=.32$, $p=.74$, $\beta=.001$. For performance avoidance goals, there was a small but significant Intervention × Prior achievement interaction, in the direction that students with higher levels of prior achievement benefitted slightly more, $t(6631)=-2.23$, $p=.03$, $\beta=.03$.

Study 2: Does a Re-designed Intervention Improve Grades?

In Study 1, the revised growth mindset intervention outperformed its predecessor in terms of changes in immediate self-reports and behavior. In Study 2 we examined whether this revised intervention would improve actual grades among 9th graders just beginning high school and replicate the effects of prior studies.

We carried out this experiment with a census (>95%) of students in 10 schools. In addition, instead of conducting the experiment ourselves (as in Study 1 and prior research), we contracted a third-party research firm specializing in government-sponsored public health surveys to collect and clean all data.

These procedural improvements have scientific value. First, in prior research that served as the basis for the present investigation (Paunesku et al, 2015), the average proportion of students in the high school who completed the Time 1 session was 17%. The goal of that research (and Study 1) was to achieve sample size, not within-school representativeness.[5] Thus, the present study may include more of the kinds of students that may have been underrepresented in prior studies.

Next, achieving a census of students is informative for policy. As noted at the outset, schools, districts, and states are often interested in raising the achievement for entire schools or

[5] In Study 1 we were not able to estimate the % of students within each school who participated because schools did not provide any official data for the study. Yet using historical numbers from the common core of data as a denominator for a sub-sample of schools that could be matched, the median response rate was approximately 40%.

for entire defined sub-groups, not for groups of students whose teachers or schools may have selected them into the experiment on the basis of their likelihood of being affected by the intervention.

Finally, ambiguity about seemingly mundane methodological choices is one important source of the non-replication of psychological experiments (see, e.g., Schooler, 2014). Therefore it is important for replication purposes to be able to train third-party researchers in study procedures, and have an arms-length relationship with data cleaning, which was done here.

Method

Data. Participants were a maximum of 3,676 students from a national convenience sample of ten schools in California, New York, Texas, Virginia, and North Carolina. One additional school was recruited, but validated student achievement records could not be obtained. The schools were selected from a national sampling frame based on the Common Core of Data, with these criteria: public high school, 9^{th} grade enrollment between 100 and 600 students, within the medium range for poverty indicators (e.g. free or reduce price lunch %), and moderate representation of students of color (Hispanic / Latino or Black / African-American). Schools selected from the sampling frame were then recruited by a third party firm. School characteristics—e.g., demographics, achievement, and average fixed mindset score—are in Table 4.

Student participants were somewhat more diverse than Study 1: 29% were Hispanic/Latino, 17% were black/African-American, 3% were Native American / American Indian, 30% were White, non-Hispanic, 6% were Asian / Asian-American, and the rest were from another or multiple racial groups. Forty-eight percent were female, and 52% reported that their mothers had earned a Bachelor's degree or greater.

The response rate for all eligible 9th grade students in the 10 participating schools for the Time 1 session was 96%. A total of 183 students did not enter their names accurately and were not matched at Time 2. All of these unmatched students received control exercises at Time 2. An additional 291 students completed Time 1 materials but not Time 2 materials, and this did not vary by condition (Control = 148, Intervention = 143). There were no data exclusions. Students were retained as long as they began the Time 1 survey, regardless of Time 2 participation or quality of responses—that is, we estimated "intent-to-treat" (ITT) effects. ITT effects are conservative tests of the hypothesis, they afford greater internal validity (preventing possible differential attrition from affecting results), and they are more policy-relevant because they demonstrate the effect of offering an intervention, which is what institutions can control.

Procedures. The firm collected all data directly from the school partners, and cleaned and merged it, without influence of researchers. Before the final dataset was delivered, the research team pre-registered the primary hypothesis and analytic approach via the Open Science Framework (OSF). The pre-registered hypothesis, a replication of Paunesku et al's (2015) interaction effect, was that *prior achievement, indexed by a composite of 8th grade GPA and test scores, would moderate mindset intervention effects on 9th grade core course GPA and D/F averages* (see: osf.io/aerpt; deviations from the pre-analysis plan are disclosed in Appendix 1).

Intervention delivery. Student participation consisted of two one-period online sessions conducted at the school, during regular class periods, in a school computer lab or classroom. The sessions were 1-4 weeks apart, beginning in the first 10 weeks of the school year. Sessions consisted of survey questions and the intervention or control intervention. Students were randomly assigned by the software, in real time, to the intervention or control group. A script was read to students by their teachers at the start of each computer session.

Mindset intervention. This was identical to the revised mindset intervention in Study 1.

Control activity. The control activity was designed to be parallel to the intervention activity. It, too, was framed as providing helpful information about the transition to high school, and participants were asked to read and retain this information so as to write their opinions and help future students. Because the revised mindset intervention had been shown to be more interesting that previous versions, great effort was made to make the control group at least as interesting as the intervention.

The control activity involved the same type of graphic art (e.g., images of the brain, animations), as well as compelling stories (e.g., about Phineas Gage). It taught basic information about the brain, which might have been useful to students taking 9th grade biology. It also provided stories from upperclassmen, reporting their opinions about the content. The celebrity stories and quotes in Time 2 were matched but they differed in content. For instance, in the control activity Michelle Obama talked about the White House's BRAIN initiative, an investment in neuroscience. Finally, as in the intervention, there were a number of opportunities for interactivity; students were asked open-ended questions and they provided their reactions.

Measures.

9th grade GPA. Final grades for the end of the first semester of 9th grade were collected. Schools that provided grades on a 0-100 scale were asked which level of performance corresponded to which letter grade (A+ to F), and letter grades were then converted to a 0 to 4.33 scale. Using full course names from transcripts, researchers, blind to students' condition or grades, coded the courses as science, math or English (i.e., core courses) or not. End of term grades for the core subjects were averaged. When a student was enrolled in more than one course

in a given subject (e.g., both Algebra and Geometry), the student's grades in both were averaged, and then the composite was averaged into their final grade variable.

As a second measure using the same outcome data, we created a dichotomous variable to indicate poor performance (1 = an average GPA of D+ or below, 0 = not; this dichotomization cut-point was pre-registered: osf.io/aerpt).

Prior achievement. The 8th grade prior achievement variable was an unweighted average of 8th grade GPA and 8th grade state test scores, which is standard measure in prior intervention experiments with incoming 9th graders (e.g. Yeager, Johnson, et al., 2014). The high schools in the present study were from different states and taught students from different feeder schools. We therefore z-scored 8th grade GPA and state test scores. Because this removes the mean from each school, we later tested whether adding fixed effects for school to statistical models changed results (it did not; see Table 6).[6] A small proportion was missing both prior achievement measures and they were assigned a value of zero; we then included a dummy-variable indicating missing data, which increases transparency and is the prevailing recommendation in program evaluation (Puma, Olsen, Bell, & Price, 2009).

Hypothetical challenge-seeking. This measure was identical to Study 1. The make-a-worksheet task was not administered in this study because it was not yet developed when Study 2 was launched.

Fixed mindset, attributions, and performance goals. These measures were identical to Study 1.

---

[6] Some prior research (Hulleman & Harackiewicz, 2009) used self-reported expectancies, not prior GPA, as a baseline moderator. We too measured these (see Study 1 for measure). When added to the composite, our moderation results were the same and slightly stronger. We did not ultimately include this measure in our baseline composite in the main text because we did not pre-register it.

Fidelity measures. Measures of distraction, interest, and self-reported learning were the same as Study 1.

Results

Preliminary analyses.

Random assignment. Random assignment to condition was effective. There were no differences between conditions in terms of demographics (gender, race, ethnicity, special education, parental education) or in terms of prior achievement within any of the 10 schools or in the full sample (all *p*s>.1). As shown in Table 3, there were no pre-intervention differences between conditions in terms of fixed mindset.

Fidelity. Several measures suggest high fidelity of implementation. On average, 94% of treated and control students answered the open-ended questions at both Time 1 and 2, and this did not differ by conditions or time. During the Time 1 session, both treated and control students saw an average of 96% of the screens in the intervention. Among those who completed the Time 2 session, treated students saw 99% of screens, compared to 97% for control students. Thus, both conditions saw and responded to their respective content to an equal (and high) extent.

Open-ended responses from students confirm that they were, in general, processing the mindset message. Here are some examples of student responses to the final writing prompt at Time 2, in which they were asked to list the next steps they could take on their growth mindset paths:

> "I can always tell myself that mistakes are evidence of learning. I will always find difficult courses to take them. I'll ask for help when I need it."

> "To get a positive growth in mindset, you should always ask questions and be curious. Don't ever feel like there is a limit to your knowledge and when feeling stuck, take a few deep breaths and relax. Nothing is easy."

> "Step 1: Erase the phrase 'I give up' and all similar phrases from your vocabulary. Step 2: Enter your hardest class of the day (math, for example). Step 3: When presented with a brain-frying worksheet, ask questions about what you don't know."

> "When I grow up I want to be a dentist which involves science. I am not good at science…yet. So I am going to have to pay more attention in class and be more focused and pay attention and learn in class instead of fooling around."

Next, students across the two conditions reported no differences in levels of distraction, own distraction: $t(3438)=0.15$, $p=.88$; others' distraction: $t(3438)=0.37$, $p=.71$. Finally, control participants actually rated their content as more interesting, and said that they felt like they learned more, as compared to treated participants, $t(3438)=7.76$ and 8.26, Cohen's ds = .25 and .27, respectively, ps <.001. This was surprising but it does not threaten our primary inferences. Instead it points to the conservative nature of the control group. Control students received a positive, interesting, informative experience that held their attention and exposed them to novel scientific information relevant to their high school biology classes (e.g., the brain) that was endorsed by influential role models (e.g., Michele Obama, LeBron James).

Self-reported fixed mindset. As an additional manipulation check, students reported their fixed mindset beliefs. As shown in Table 3, both the intervention and control conditions changed

in the direction of a growth mindset between Times 1 and 2 (change score *p*s<.001). However, those in the intervention condition changed much more (Control Δ = -.17 scale points out of 6, Intervention Δ = -.55 scale points out of 6). These change scores differed significantly from each other, *p*<.001. Table 5 reports regressions predicting post-intervention fixed mindset as a function of condition, and shows that choices of covariates did not affect this result.

Moderator analyses in Table 5 show that previously higher-achieving students, and, to a much lesser extent, students who held more of a fixed mindset at baseline, changed more in the direction of a growth mindset.

It is interesting that the control condition showed a significant change in growth mindset—almost identical to the original mindset condition in Study 1. Perhaps teachers were regularly discussing growth mindset concepts, perhaps treated students behaved in a more growth-mindset-oriented way, spilling over to control students, or perhaps the control condition itself—by creating strong interest in the science of the brain and making students feel as though they learned something—implicitly taught a growth mindset. In any of these cases, this would make the treatment effect on grades conservative.

Primary analyses.

***9th grade GPA and poor performance rates*.** Our first pre-registered confirmatory analysis was to examine the effects of the intervention on 9th grade GPA, moderated by prior achievement. In the full sample, there was a significant Intervention × Prior Achievement interaction, $t(3419)=2.66$, $p=.007$, $\beta=-.05$, replicating prior research (Paunesku et al., 2015; Yeager et al., 2014; also see Wilson & Linville, 1982; 1985). Table 6 shows that the significance of this result did not depend on the covariates selected. Tests at ±1SD of prior performance showed an estimated intervention benefit of 0.13 grade points, $t(3419)=2.90$, $p=.003$, $d=.10$, for

those who were at -1SD of prior performance, and no effect among those at +1SD of prior performance, b=-0.04 grade points, $t(3419)=0.99$, p=.33, d=.03. This may be because higher achieving students have less room to improve, or because they may manifest their increased growth mindset in challenging-seeking rather than in seeking easy A's.

In a second pre-registered confirmatory analysis, we analyzed rates of poor performance (D or F averages). This analysis mirrors past research (Cohen et al., 2009; Paunesku et al., 2015; Wilson & Linville, 1982, 1985; Yeager, Purdie-Vaughns, et al., 2014), and helps test the theoretically predicted finding that the intervention is beneficial by stopping a recursive process by which poor performance begets worse performance over time (Cohen et al., 2009; see Cohen & Sherman, 2014; Yeager & Walton, 2011).

There was a significant overall main effect of intervention on a reduced rate of poor performance of 4 percentage points, Z=2.95, p=.003, d=.10. Next, as with the full continuous GPA metric, in a logistic regression predicting poor performance there was a significant Intervention × Prior Achievement interaction, Z=2.45, p=.014, β=.05. At -1SD of prior achievement the intervention effect was estimated to be 7 percentage points, Z=3.80, p<.001, d=.13, while at +1SD there was a non-significant difference of 0.7 percentage points, Z=.42, p=.67, d=.01.

Secondary analyses.

Hypothetical challenge-seeking. The mindset intervention reduced from 54% to 45% the proportion of students saying they would choose the "easy" math homework assignment (that they would likely get a high score on) versus the "hard" assignment (that they might get a low score on) (see Table 3). The intervention effect on hypothetical challenge-seeking was slightly larger for previously higher-achieving students, Intervention × Prior Achievement interaction

Z=2.63, p=.008, β=.05, and was not moderated by pre-intervention fixed mindset, Z=1.45, p=.15, β=.02. Thus, while lower achieving students were more likely than high achieving students to show benefits in grades, higher achieving students were more likely to show an impact on their challenge-seeking choices on the hypothetical task.

***Attributions and goals*.** The growth mindset intervention reduced fixed-trait, person-focused attributions, d=.13, and performance avoidance goals, d=.11, ps<.001 (see Table 3), unlike some prior growth mindset intervention research, which did not change these measures (Blackwell et al., 2007, Study 2). Thus, the present study uses a field experiment to replicate much prior research (Burnett et al., 2013; Dweck, 1999; Dweck, 2006; Yeager & Dweck, 2012).

General Discussion

The present research used "design thinking" to make psychological intervention materials more broadly applicable to students who may share concerns and construals because they are undergoing similar challenges—in this case, 9th grade students entering high school. When this was done, the revised intervention was more effective in changing proxy outcomes such as beliefs and short-term behaviors than previous materials (Study 1). Furthermore, the intervention increased core course grades for previously low-achieving students (Study 2).

Although we do not consider the revised version to be the final iteration, the present research provides direct evidence of an exciting possibility: a two-session mindset program, developed through an iterative, user-centered design process, may be administered to entire classes of 9th graders (>95% of students) and begin to raise the grades of the lowest performers, while increasing the learning-oriented attitudes and beliefs of low and high performers. This approach illustrates an important step toward taking growth mindset and other psychological interventions to scale, and for conducting replication studies.

At a theoretical level, it is interesting to note the parallels between mindset interventions and expert tutors: both are strongly focused on students' construals (see Lepper, Woolverton, Mumme, & Gurtner, 1993; Treisman, 1992). Working hand-in-hand with students, expert tutors build relationships of trust with students and then redirect their construals of academic difficulty as challenges to be met, not evidence of fixed inability. In this sense, both expert tutors and mindset interventions recognize the power of construal-based motivational factors in students' learning. Future interventions might do well to capitalize further on the wealth of knowledge of expert tutors, who constitute one of the most powerful educational interventions (Bloom, 1984).

Replication

Replication efforts are important for cumulative science (Funder et al., 2014; Lehrer, 2010; Open Science Collaboration, 2015; Schooler, 2011, 2014; Schimmack, 2012; Simmons et al., 2011; Pashler & Wagenmakers, 2012; also see Ioannidis, 2005). However, it would be easy for replications to take a misguided "magic bullet" approach—that is, to assume that intervention materials and procedural scripts that worked in one place for one group should work in another place for another group (Yeager & Walton, 2011). This is why Yeager and Walton (2011) stated that experimenters "should [not] hand out the original materials without considering whether they would convey the intended meaning" for the group in question (p. 291; also see Wilson, Aronson, & Carlsmith, 2010).

The present research therefore followed a procedure of (a) beginning with the original materials as a starting point; (b) using a design methodology to increase the likelihood that they conveyed the intended meaning in the targeted population (9^{th} graders); and (c) conducting well-powered randomized experiments in advance to ensure that, in fact, the materials were appropriate and effective in the target population.

Study 2 showed that, when this was done, the revised mindset intervention raised grades for previously low-achieving students, when all data were collected, aggregated, and cleaned by an independent third-party, and when all involved were blind to experimental condition. Furthermore, the focal tests were pre-registered (though see Appendix 1). The present research therefore provides a rigorous replication of Paunesku et al. (2015). This directly addresses skepticism about whether a two-session, self-administered psychological intervention can, under some conditions, measurably improve the grades for previously low-performing students.

Nevertheless, the pre-registered hypothesis reported here does not advance our theory of mechanisms for psychological intervention effects. Additional analyses of the data—ones that could not have been pre-registered—will be essential for doing this and further improving the intervention and its impact.

Divergence of Self-Reports and Behavior

One theme in the present study's results—and a theme dating to the original psychological interventions in education (Wilson & Linville, 1982)—is a disconnect between self-reports and actual behavior. On the basis of only the self-reported fixed mindset results, we might have concluded that the intervention benefitted high achievers more. However, on the basis of the GPA results, we might conclude that the intervention "worked" better for students who were lower in prior achievement. That is, self-reports and grades showed moderation by prior performance in the *opposite* directions.

What accounts for this? Two possibilities are germane. First, grades follow a non-normal distribution that suffers from a range restriction at the top, in part due to grade inflation. Thus higher-achievers may not have room to increase their grades. But higher-achievers may also be

those who read material more carefully and absorb the messages they are taught. These patterns could explain smaller GPA effects but larger proxy outcome effects for higher-achievers.

Another possibility is that high-achieving students may have been inspired to take on more challenging work that might not lead to higher grades but might teach them more. Indeed, the hypothetical challenge-seeking measure supports this. For this reason, it might have been informative to analyze scores on a norm-referenced exam at the end of the term, as in some intervention studies (Good et al., 2003; Hanselman, Bruch, Gamoran, & Borman, 2014). If higher-achievers were choosing harder tasks, their grades might have suffered, but their learning might have improved. In the present research it was not possible to administer a norm-referenced test due to the light-touch nature of the intervention. In addition, curriculum was not held constant across our diverse sample of schools.

Finally, Study 1's development of a brief behavioral measure of challenge seeking—the "make-a-worksheet" task—was a methodological advance that may help avoid known problems with using self-reports to evaluate behavioral interventions (see Duckworth & Yeager, 2015). This may make it a practical proxy outcome in future evaluation of interventions (see Duckworth & Yeager, 2015; Yeager & Bryk, 2015).

Limitations

The present research has many limitations. First, we only had access to grades at the end of the first term of 9th grade—in some cases, a few weeks after the intervention. It would have been informative to collect grades over a longer period of time.

Second, we have not attempted here to understand heterogeneity in effect sizes (cf. Hanselman et al. 2014). Building on the results of the present study, we are now poised to carry out a systematic examination of student, teacher, and school factors that cause heterogeneous

intervention effects. Indeed, with further revision to the intervention materials, we are carrying out a version of Study 2 with students in a national sample of randomly-selected U.S. public high schools. The present study provides a basis for this.

Third, it is interesting that, although the revised mindset intervention was more effective at changing short-term beliefs and behaviors than the original intervention (Study 1), the effect sizes for grades in Study 2 do not exceed those found in past analogous studies (e.g. Paunesku et al., 2015; Yeager, Henderson, et al., 2014). Perhaps the present study, in reaching a census of students, included more students who were reluctant to participate or who may have been resistant to the intervention message. We furthermore utilized intent-to-treat analyses—including students who saw any content at session 1 regardless of whether they completed session 2. Or perhaps the present study enlisted teachers who were less compliant with procedures than the enthusiastic early adopter teachers recruited in past studies. With lower experimental control and a census of students, there may be greater error variance, resulting in lower effect sizes. Finally, we only examined fall term grades, many of which were measured just weeks after the treatment.

Although the mindset effect sizes may appear to be small—and they are compared to the magnitude of student engagement issues in U.S. schools—they they conform to expectations,[7] and are nevertheless practically meaningful. Considering that 9th grade course failure almost perfectly predicts eventual high school graduation rates (Allensworth & Easton, 2015), if the growth mindset intervention reduces by 4 percentage points the proportion of 9th graders who earn D/F averages, then a fully scaled and spread version of this program could in theory prevent

[7] The correlations with fixed mindset beliefs in Table 2 ranged from r=.12 to r=.40, which matches a meta-analysis of implicit theories effects (range of r=.15 to r=.24; Burnette et al., 2013). The present effects also match to the average of social-psychological effects more generally, r=.21 (Richard, Bond, & Stokes-Zoota, 2003).

100,000 high school dropouts in the U.S. per year—while increasing the learning-oriented behavior of many other students.

Conclusion

This research began with a simple question: is it possible to take an existing, initially effective psychological intervention and re-design it to improve the outcomes for a population of students undergoing a similar life transition? Now that we have shown this is possible for a growth mindset intervention and now that we have described a method for doing so, it will be exciting to see how this approach can be applied to other interventions. If they too can be improved, would it be possible to disseminate them in such a way that would actually reduce the large number of high school dropouts for the nation as a whole? We look forward to investigating this.

References

Allensworth, E. M., & Easton, J. Q. (2005). *The on-track indicator as a predictor of high school graduation*. Consortium on Chicago School Research, University of Chicago.

Aronson, E. (1999). The power of self-persuasion. *American Psychologist, 54,* 875–884.

Aronson, J., Fried, C., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38,* 113–125.

Bem, D. J. (1965). An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology, 1,* 199-218.

Blackwell, L. A., Trzesniewski, K. H., & Dweck, C. S. (2007). Theories of intelligence and achievement across the junior high school transition: A longitudinal study and an intervention. *Child Development, 78,* 246–263.

Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality, 13,* 420-432.

Brehm, J. W. (1966). *A theory of psychological reactance.* New York, NY: Academic Press.

Bryk, A. S. (2009). Support a science of performance improvement. *Phi Delta Kappan, 90(8),* 597-600.

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

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and re-evaluation. *Advances in Experimental Social Psychology, 24,* 201–234.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313,* 1307–1310.

Cohen, G. L. & Prinstein, M. J. (2006). Peer contagion of aggression and health-risk behavior among adolescent males: An experimental investigation of effects on public conduct and private attitudes. *Child Development, 77*, 967-983.

Cohen, G. L., Garcia, J., Purdie-Vaugns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science, 324*, 400–403.

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology, 65,* 333-371.

Cooper, J., & Fazio, R. H. (1984). A new look at dissonance theory. *Advances in Experimental Social Psychology, 17,* 229-266.

Davis, J., & Bauman, K. (2013). School enrollment in the United States: 2011. *United States Census Bureau*.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment, 91,* 166-174.

Duckworth, A.L., & Yeager, D.S. (in press). Measurement matters: Assessing personal qualities other than cognitive ability. *Educational Researcher*.

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

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

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

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53, 109*-132.

Elliot, A. J., & Dweck, C. S. (Eds.) (2005). *Handbook of competence and motivation.* New York, NY: Guilford Press.

Erikson, E. H. (1968). *Identity: Youth and Crisis.* New York, NY: Norton.

Fryberg, S. A., Covarrubias, R., & Burack, J. (2013). Cultural models of education and academic performance for Native American and European American students, *School Psychology International, 34,* 439-452.

Funder, D. C., Levine, J. M., Mackie, D. M., Morf, C. C., Sansone, C., Vazire, S., & West, S. G. (2014). Improving the dependability of research in personality and social psychology: Recommendations for research and educational practice. *Personality And Social Psychology Review, 18,* 3-12.

Garcia, J., & Cohen, G. L. (2012). Social psychology and educational intervention. In E. Shafir (Ed.), *Behavioral foundations of policy.* New York, NY: Russell Sage.

Godes, O., Hulleman, C. S., & Harackiewicz, J. M. (2007). Boosting students' interest in math with utility value: Two experimental tests. *Meeting of the American Educational Research Association, Chicago, IL.*

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24,* 645–662.

Hanselman, P., Bruch, S.K., Gamoran, A., & Borman, G.D. (2014). Threat in context: School

moderation of the impact of social identity threat on racial/ethnic achievement gaps. *Sociology of Education, 87,* 106-124.

Hasebe, Y., Nucci, L., & Nucci, M. S. (2004). Parental control of the personal domain and adolescent symptoms of psychopathology: A cross-national study in the United States and Japan. *Child Development, 75,* 815-828.

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

Hulleman, C. S., & Harackiewicz, J. M. (2009). Making education relevant: Increasing interest and performance in high school science classes. *Science, 326,* 1410–1412.

Ioannidis, J. P. (2005). Why most published research findings are false. *Plos One, 18,* 40-47.

Kelley, T., & Kelley, D. (2013). *Creative confidence: Unleashing the creative potential within us all.*

Kohavi, R., & Longbotham, R. (2015). Online controlled experiments and A/B tests. In C. Sammut and G. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining*. NY, NY: Springer.

Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75,* 63-82.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory.* NY, NY: Oxford University Press.

Lehrer, J. (2010). The truth wears off. *The New Yorker, 13,* 52.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. L. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools: Technology in education*. Hillsdale, NJ: Erlbaum.

Lepper, M. R., & Woolverton, M. (2001). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Contributions of social psychology* (pp.133-156). Orlando, FL: Academic Press.

Lewin, K. (1952). Group decision and social change. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (2nd ed., pp. 330–344). New York, NY: Holt.

Lord, C. G. (1980). Schemas and images as memory aids: Two modes of processing social information. *Journal of Personality and Social Psychology, 38*, 257-269.

McCannon, C.J., Schall, M.W., & Perla, R.J. (2008). *Planning for Scale: A Guide for Designing Large-Scale Improvement Initiatives*. *IHI Innovation Series white paper.* Cambridge, Massachusetts: Institute for Healthcare Improvement.

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

Nucci, L. P., Killen, M., & Smetana, J. G. (1996). Autonomy and the personal: Negotiation and social reciprocity in adult-child social exchanges. *New Directions for Child and Adolescent Development, 1996,* 7-24.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349,* 943.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science a crisis of confidence?. *Perspectives on Psychological Science, 7,* 528-530.

Paunesku, D., Walton, G.M., Romero, C.L., Smith, E.N., Yeager, D.S., & Dweck, C.S. (2015). Mindset interventions are a scalable intervention for academic underachievement. *Psychological Science, 26,* 784-793.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials (NCEE 2009-0049).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Richard, F.D., Bond, C.F., & Stokes-Zoota, J.J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7,* 331-363.

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Random House LLC.

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,* 313-336.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York, NY: McGraw-Hill.

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17,* 551.

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature, 470,* 437.

Schooler, J.W. (2014). Turning the lens of science on itself: Verbal overshadowing, replication,

and metascience. *Perspectives on Psychological Science, 9,* 579-584.

Simmons, J.P., Nelson, L.D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359-1366.

Skurnik, I., Yoon, C. Park, D.C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31,* 713-724.

Stephens, N.M., Hamedani, M.G., & Destin, M. (2014). Closing the social-class achievement gap: A difference-education intervention improves first-generation students' academic performance and all students' college transition. *Psychological Science, 25,* 943-953.

Stipek, D. (2002). *Motivation to learn: From theory to practice*. (4th edition). Allyn & Bacon, Needham Heights, MA.

Treisman, U. (1992). Studying students studying calculus: A look at the lives of minority mathematicians. A Mary P. Dolciani Lecture. *College Mathematics Journal, 23,* 362-372.

Tsukayama, E., Duckworth, A. L., & Kim, B. E. (2013). Domain-specific impulsivity in school-age children. *Developmental Science, 16,* 879-893.

Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science, 23,* 73-82.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92,* 82-96.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes among minority students. *Science, 331,* 1447–1451.

Wilson, T.D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, M.A.: Harvard University Press.

Wilson, T. D., Aronson, E., & Carlsmith, K. (2010). Experimentation in social psychology. *Handbook of Social Psychology* (pp. 51-81).

Wilson, T. D., & Linville, P. W. (1982). Improving the academic performance of college freshmen: Attribution therapy revisited. *Journal of Personality and Social Psychology, 42*, 367–376.

Wilson, T. D., & Linville, P. W. (1985). Improving the performance of college freshmen with attributional techniques. *Journal of Personality and Social Psychology, 49,* 287–293.

Yeager, D.S., & Bryk, A. (2015). Practical measurement. *Unpublished Manuscript, University of Texas at Austin.*

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

Yeager, D.S., Henderson, M., Paunesku, D., Walton, G., Spitzer, B., D'Mello, S., & Duckworth, A.L. (2014). Boring but important: A self-transcendent purpose for learning fosters academic self-regulation. *Journal of Personality and Social Psychology, 107,* 559-580.

Yeager, D.S., Johnson, R., Spitzer, B., Trzesniewski, K., Powers, J., & Dweck, C.S. (2014). The far-reaching effects of believing people can change: Implicit theories of personality shape stress, health, and achievement during adolescence. *Journal of Personality and Social Psychology, 106,* 867-884.

Yeager, D.S., Purdie-Vaughns, V., Garcia, J., Apfel, N., Pebley, P., Master, A., Hessert, W., Williams, M. & Cohen, G.L. (2014). Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. *Journal of Experimental Psychology: General, 143,* 804-824.

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

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

Table 1. Designing the Revised Mindset Intervention: Manipulations and Results from Rapid, A/B Experiments of Growth Mindset Intervention Elements Conducted on MTurk.

A/B Study	Total *N*	Manipulation and coding	Sample text	Effect on pre-post mindset Δ β=	*p*=
1	1851				
		Direct =1 (this will help you) framing vs. Indirect = 0 (help other people) framing	Direct: "Would you like to be smarter? Being smarter helps teens become the person they want to be in life… In this program, we share the research on how people can get smarter." vs. Indirect: "Students often do a great job explaining ideas to their peers because they see the world in similar ways. On the following pages, you will read some scientific findings about the human brain. ... We would like your help to explain this information in more personal ways that students will be able to understand. We'll use what we learn to help us improve the way we talk about these ideas with students in the future."	-0.234	0.056
		Refuting fixed mindset =1 vs. Not = 0	Refutation: Some people seem to learn more quickly than others, for example, in math. You may think, "Oh, they're just smarter." But you don't realize that they may be working really hard at it (harder than you think).	-0.402	0.002
		Labeling and explaining benefits of "growth mindset" = 1 vs. Not = 0	Benefits of mindset: People with a growth mindset know that mistakes and setbacks are opportunities to learn. They know that their brains can grow the most when they do something difficult and make mistakes. We studied all the 10th graders in the nation of Chile. The students who had a growth mindset were 3 times more likely to score in the top 20% of their class. Those with a fixed mindset were more likely to score in the bottom 20%	0.357	0.006
		Rats/jugglers scientific evidence = 1 vs. Teenagers = 0	Rats/Jugglers: Scientists used a brain scanner (it's like a camera that looks into your brain) to compare the brains of the two groups of people. They found that the people who learned how to juggle actually grew the parts of their brains that control juggling skills vs. Teenagers: Many of the [teenagers] in the study showed large changes in their intelligence scores. And these same students also showed big changes in their brain. ... This shows that teenagers' brains can literally change and become smarter—if you know how to make it happen.	-0.122	0.352
2	1153				
		Direct (this will help you) framing = 1 vs. Indirect (help other people) framing = 0	Direct: Why does getting smarter matter? Because when people get smarter, they become more capable of doing the things they care about. Not only can they earn higher grades and get better jobs, they can have a bigger impact on the world and on the people they care about... In this program, you'll learn what science says about the brain and about making it smarter. Vs. Indirect: (see above).	-0.319	0.036
		Refuting fixed mindset = 1 vs. Not=0	Refutation: Some people look around and say "How come school is so easy for them, but I have to work hard? Are they smarter than me?" ... They key is not to focus on whether you're smarter than other people. Instead, focus on whether you're smarter today than you were yesterday and how you can get smarter tomorrow than you are today.	0.002	0.990
		Celebrity endorsements = 1 vs. Not=0	Celebrity: Endorsements from Scott Forstall, LeBron James, and Michelle Obama.	0.273	0.073

Note: β= standardized regression coefficient for condition contrast in multiple linear regression.

Table 2. Correlations Among Measures in Study 1 Replicate Prior Growth Mindset Effects and Validate the "Make-A-Worksheet" Task.

	Actual Easy (Minus Hard) Problems Selected	Grit	Self-control	Fixed Mindset (Time 2)	Fixed trait attributions	Performance avoidance goals	Prior Performance	Interest in Math	Math Anxiety
Grit	-.16								
Self-control	-.14	.51							
Fixed Mindset (Time 2)	.13	-.16	-.16						
Prior Performance	-.17	.40	.37	-.25					
Fixed trait attributions	.17	-.27	-.22	.28	-.30				
Performance avoidance goals	.10	-.13	-.14	.23	-.14	.21			
Interest in Math	-.23	.27	.29	-.14	.48	-.27	-.10		
Math Anxiety	.11	-.07	-.08	.12	-.35	.23	.14	-.29	
Hypothetical willingness to select the easy (not hard) math problem.	.29	-.19	-.16	.12	-.14	.25	.12	-.26	.12

Note: *N*s range from 6,883 to 7,251; all *p*s < .01. Data are from both conditions; correlations did not differ across experimental conditions.

Table 3. Effects of Condition on Fixed Mindset, Attributions, Performance-avoidance Goals, and Challenge-Seeking in Studies 1 and 2.

	Study 1					Study 2				
	Original Mindset Intervention		Revised Mindset Intervention			Placebo Control		Revised Mindset Intervention		
	M	*SD*	*M*	*SD*	Comparison	*M*	*SD*	*M*	*SD*	Comparison
Pre-intervention (Time 1) fixed mindset	3.20	1.15	3.22	1.15	*t*=1.18	3.07	1.12	3.09	1.14	*t*=0.30
Post-intervention (Time 2) fixed mindset	2.98	1.21	2.74	1.21	*t*=9.32***	2.89	1.14	2.54	1.16	*t*=12.16***
Change from Time 1 to Time 2	-0.22		-0.48			-0.17		-0.55		
Fixed trait attributions	2.14	0.87	2.08	0.82	*t*=2.71**	2.12	0.86	2.03	0.82	*t*=3.58***
Performance avoidance goals	3.40	1.51	3.31	1.52	*t*=2.60**	3.57	1.55	3.42	1.56	*t*=2.95**
Actual easy (minus hard) math problems selected at Time 2	4.62	11.52	2.42	11.38		-	-	-	-	
Hypothetical willingness to select the easy (not hard) math problem at Time 2	60.4%		51.2%		*t*=7.95***	54.4%		45.3%		*t*=5.71***
N (both Time 1 and 2)=	3665		3480			1646		1630		

Note: Mindset change scores from Time 1 to Time 2 significant at $p<.001$. *** $p<.001$, ** $p<.01$.

Table 4. School Characteristics in Study 2.

School	School year start date	Modal start date for Time 1	Greatschools.org rating	Average pre-intervention fixed mindset	% White	% Hispanic / Latino	% Black / African-American	% Asian	% living below poverty line (in district)	State
1	8/18/14	10/6/14	4	3.13	20	26	32	22	18.3	CA
2	8/18/14	10/3/14	2	3.22	3	57	33	7	18.3	CA
3	9/3/14	10/23/14	2	3.23	10	13	74	3	41.0	NY
4	8/25/14	9/30/14	7	2.65	62	19	10	9	10.8	NC
5	8/25/14	10/6/14	5	3.11	41	17	40	1	13.7	NC
6	8/25/14	10/7/14	2	2.95	29	19	48	2	13.7	NC
7	8/26/14	9/23/14	7	3.00	78	21	1	0	11.6	TX
8	8/25/14	9/18/14	4	3.21	11	78	6	4	27.8	TX
9	8/25/14	10/8/14	8	3.07	52	27	11	9	27.8	TX
10	9/2/14	10/29/14	6	3.43	55	16	26	3	5.9	VA

Table 5. Regressions Predicting Post-intervention (Time 2) Fixed Mindset, Study 2

	Base model	Plus school fixed effects	Plus demographic covariates	Plus pre-intervention mindset
Intercept	2.959***	2.933***	2.943***	2.989***
	(0.028)	(0.079)	(0.086)	(0.071)
Revised mindset intervention	-0.389***	-0.394***	-0.397***	-0.391***
	(0.039)	(0.039)	(0.039)	(0.032)
Prior achievement (z-scored, centered at 0)	-0.156***	-0.182***	-0.173***	-0.036
	(0.028)	(0.029)	(0.029)	(0.025)
Intervention × Prior achievement (z-scored, centered at 0)	-0.181***	-0.177***	-0.176***	-0.183***
	(0.040)	(0.039)	(0.039)	(0.033)
Female			0.020	-0.031
			(0.039)	(0.032)
Asian			-0.115	-0.150*
			(0.084)	(0.069)
Hispanic/Latino			-0.024	-0.000
			(0.051)	(0.043)
Black/African-American			0.036	-0.008
			(0.058)	(0.048)
Repeating freshman year			0.356**	0.180
			(0.137)	(0.114)
Pre-intervention fixed mindset (z-scored, centered at 0)				0.704***
				(0.023)
Intervention × Pre-intervention fixed mindset (z-scored, centered at 0)				-0.120***
				(0.033)
Adjusted R2	.076	.099	.100	.383
AIC	10103.731	10030.747	10015.772	8759.446
N	3279	3279	3274	3267

Note: OLS regressions. Unstandardized coefficients above standard errors (in parentheses). School fixed effects included in model but suppressed from regression table.

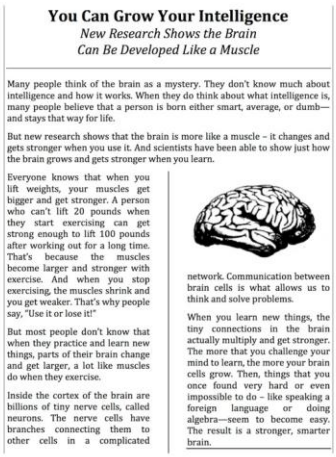
Table 6. Regressions Predicting End-of-Term GPA in Math, Science, and English, Study 2.

	Base model	Plus school fixed effects	Plus demographic covariates	Plus pre-intervention mindset
Intercept	1.556***	1.584***	1.581***	1.589***
	(0.035)	(0.063)	(0.067)	(0.066)
Revised mindset intervention (among low prior-achievers, -1SD)	0.119*	0.125**	0.123**	0.135**
	(0.049)	(0.047)	(0.045)	(0.046)
Prior achievement (z-scored, centered at -1SD)	0.693***	0.721***	0.663***	0.641***
	(0.024)	(0.024)	(0.023)	(0.024)
Intervention × Prior achievement (z-scored, centered at -1SD)	-0.079*	-0.082*	-0.080*	-0.091**
	(0.034)	(0.033)	(0.032)	(0.032)
Female			0.327***	0.338***
			(0.031)	(0.031)
Asian			0.189**	0.190**
			(0.067)	(0.067)
Hispanic/Latino			-0.287***	-0.284***
			(0.041)	(0.041)
Black/African-American			-0.322***	-0.309***
			(0.046)	(0.046)
Repeating freshman year			-0.922***	-0.894***
			(0.108)	(0.108)
Pre-intervention fixed mindset (z-scored, centered at 0)				-0.110***
				(0.022)
Intervention × Pre-intervention fixed mindset (z-scored, centered at 0)				-0.018
				(0.032)
Adjusted R2	.295	.358	.407	.416
AIC	9781.892	9467.187	9196.462	9112.995
N	3448	3448	3448	3438

Note: OLS regressions. Unstandardized coefficients above standard errors (in parentheses). School fixed effects included in model but suppressed from regression table.

Figure 1. Screenshots of the "original" and "revised" mindset interventions

Original



You Can Grow Your Intelligence

New Research Shows the Brain Can Be Developed Like a Muscle

Many people think of the brain as a mystery. They don't know much about intelligence and how it works. When they do think about what intelligence is, many people believe that a person is born either smart, average, or dumb—and stays that way for life.

But new research shows that the brain is more like a muscle – it changes and gets stronger when you use it. And scientists have been able to show just how the brain grows and gets stronger when you learn.

Everyone knows that when you lift weights, your muscles get bigger and get stronger. A person who can't lift 20 pounds when they start exercising can get strong enough to lift 100 pounds after working out for a long time. That's because the muscles become larger and stronger with exercise. And when you stop exercising, the muscles shrink and you get weaker. That's why people say, "Use it or lose it!"

But most people don't know that when they practice and learn new things, parts of their brain change and get larger, a lot like muscles do when they exercise.

Inside the cortex of the brain are billions of tiny nerve cells, called neurons. The nerve cells have branches connecting them to other cells in a complicated network. Communication between brain cells is what allows us to think and solve problems.

When you learn new things, the tiny connections in the brain actually multiply and get stronger. The more that you challenge your mind to learn, the more your brain cells grow. Then, things that you once found very hard or even impossible to do - like speaking a foreign language or doing algebra—seem to become easy. The result is a stronger, smarter brain.

Revised

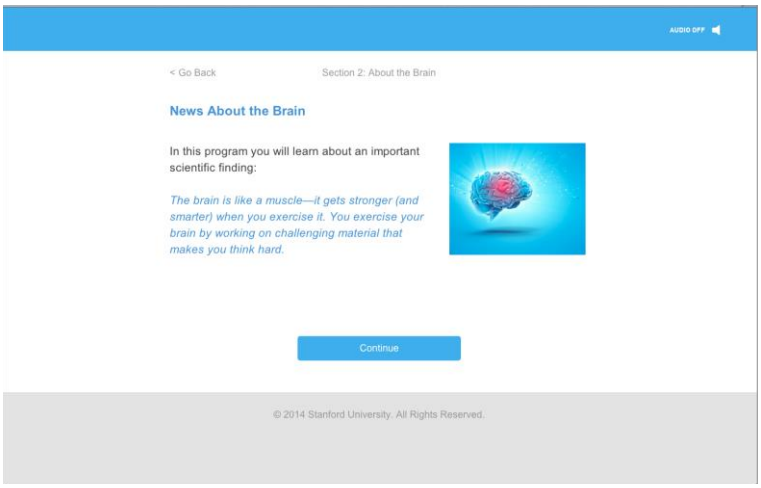
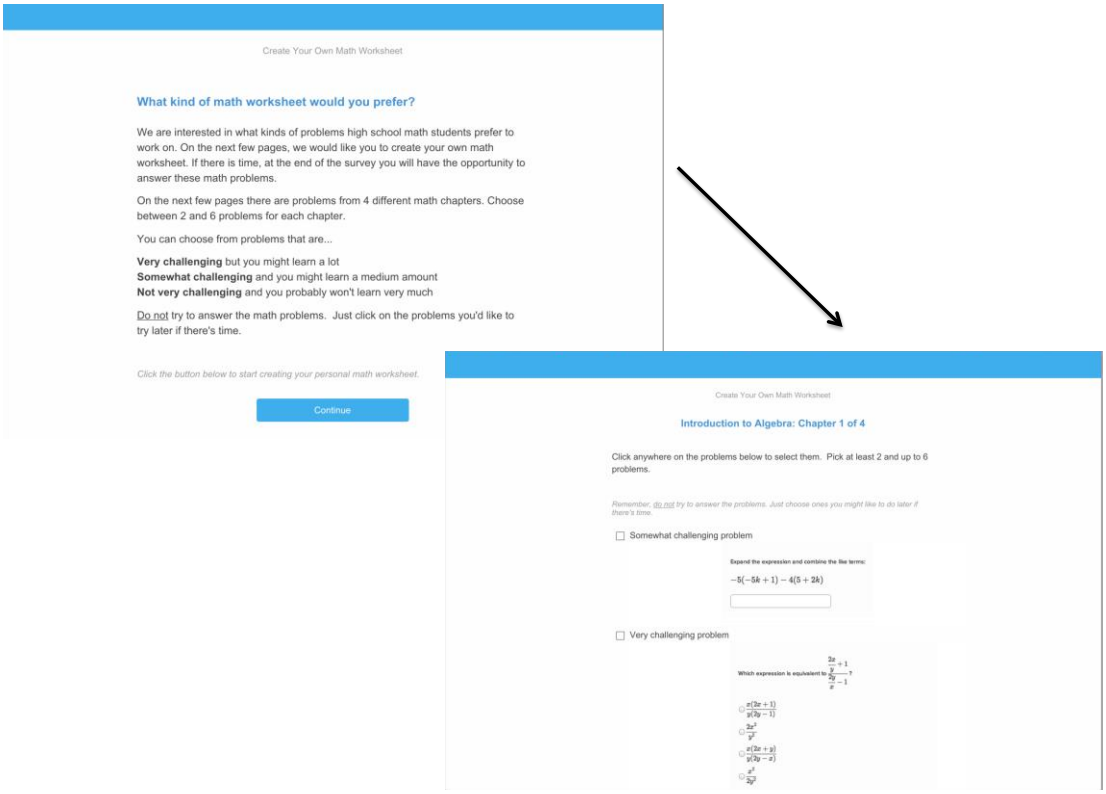


Figure 2. Screenshots from the "make a worksheet" task.



Appendix 1: Deviations from Pre-Registered Analysis Plan

The study was pre-registered before the delivery of the dataset to the researchers by the third-party research firm (osf.io/aerpt). However after viewing the data, but before testing for the effects of the intervention variable, we discovered that some of the pre-registered methods were not possible to carry out. Because at this time there are no standards for pre-registration in psychological science, the research team used its best judgment to make minor adjustments. In the end, we selected analyses that were closest to a replication of the Intervention × Prior achievement interaction reported by Paunesku et al. (2015) (Paunesku et al. also report a main effect of intervention on grades, but the primary effect that is interpreted and highlighted in the paper is the interaction). The deviations from the pre-analysis plan were:

- We initially expected to exclude students on an individualized special education plan, but that variable was not delivered to us for any schools.

- We pre-registered a second way of coding prior performance: by including self-reported prior grades in the composite. However, we later discovered a meta-analysis showing that this is inappropriate for use in tests of moderation because lower-performing students show greater bias and measurement error, making it an inappropriate moderator, even though self-reported grades are an effective linear covariate (Kuncel, Crede, & Thomas, 2005).

- We hoped to include social studies grades as a core subject, but schools did not define which classes were social studies and they could not be contacted to clarify, and so we only classified math, science and English as core classes.

- We pre-registered other analyses that will be the subject of other working papers: effects on stress (which were described as exploratory in our pre-registered plan), and moderation of intervention effects by race/ethnicity and gender (which were not found in Paunesku et al., 2015 and so, strictly speaking, they are not a part of the replication reported here).

View publication stats