Learning and Motivation 72 (2020) 101683

Contents lists available at ScienceDirect

Learning and Motivation

journal homepage: www.elsevier.com/locate/l&m

Can a brief intervention alter genetic and environmental influences on psychological traits? An experimental behavioral genetics approach

Alexander P. Burgoyne [a,*], Sarah Carroll [a], D. Angus Clark [b], David Z. Hambrick [a], Kathryn S. Plaisance [c], Kelly L. Klump [a], S. Alexandra Burt [a]

[a] Michigan State University, United States
[b] University of Michigan, United States
[c] University of Waterloo, Canada

ARTICLE INFO

Keywords:
Mindset
Intervention
Twin study
Heritability
Etiology
Experimental behavioral genetics

ABSTRACT

Mindset interventions are designed to encourage students to adopt a *growth mindset*, reflecting the belief that one's intelligence can be improved in an effort to increase academic achievement. How do these interventions exert their effects? We assessed the effects of an online mindset intervention on mindset and four outcome variables, grit, locus of control, challenge-seeking behavior, and cognitive ability test performance in a sample of 1668 twins in 840 families. The mindset intervention successfully induced a growth mindset, but had no effect on the outcome variables. Biometric modeling of self-reported mindset before and after the intervention further revealed that this mean change in mindset was accompanied by changes in its etiology. At baseline, nonshared environmental contributions to individual differences in mindset predominated, while additive genetic contributions were small-to-moderate and shared environmental contributions were negligible. After the intervention, there was an increase in additive genetic contributions to individual differences in mindset. In other words, despite its very brief nature, our simple environmental intervention acted to increase the heritability of mindset. Such findings suggest that interventions may sometimes exert their effects by altering the genetic influences on a trait.

1. Introduction

Students differ both in how readily they acquire academic skills, as well as in their ultimate level of achievement. For example, some high school students progress through advanced placement calculus courses, but others make it no further than algebra. What accounts for these individual differences in achievement? Research indicates that general cognitive ability (*g*) plays a large role. In a 5-year prospective study of over 70,000 children, *g* and academic achievement were highly correlated, $r = .81$ (Deary, Strand, Smith, & Fernandes, 2007). This very strong association has additional implications for our understanding of the etiology of academic achievement, since both *g* and academic achievement are substantially heritable, with typical heritability estimates ranging from 50 to

* Corresponding author at: School of Psychology, Georgia Institute of Technology, United States.
E-mail address: burgoyn4@gmail.com (A.P. Burgoyne).

https://doi.org/10.1016/j.lmot.2020.101683
Received 21 May 2020; Received in revised form 5 October 2020; Accepted 6 October 2020

70% (Knopik, Neiderhiser, DeFries, & Plomin, 2017; Plomin & Deary, 2015).[1] Furthermore, many of the same genetic influences that give rise to individual differences in *g* also impact academic achievement, helping to explain their strong relationship (Calvin et al., 2012).

Beyond *g*, many other factors have been proposed to contribute to individual differences in achievement. For example, *grit*, reflecting consistency of interest and perseverance of effort, is associated with educational attainment in adults and grade point average in undergraduates (Credé, Tynan, & Harms, 2017; Duckworth, Peterson, Matthews, & Kelly, 2007). Similarly, *locus of control*, reflecting whether students believe academic outcomes stem from their own behaviors, also predicts academic achievement, with small-to-moderate effects (Findley & Cooper, 1983). Finally, some have argued that *mindset* is an important predictor of academic achievement (Dweck, 2008). Students with a growth mindset believe traits such as intelligence can be improved with effort, whereas those with a fixed mindset believe these traits are stable. Growth mindset is argued to predict endorsement of learning goals, persistence to overcome challenges, and resilience following failure, culminating in greater academic achievement (Dweck & Leggett, 1988; but for recent critique of these premises, see Burgoyne, Hambrick, & Macnamara, 2020; Li & Bates, 2019). What's more, because mindset interventions are typically brief, inexpensive, and administered via computer, they have featured prominently in recent efforts to boost students' achievement.

This enthusiasm may have been premature, however. Empirical analyses have provided some support for mindset effects, but also suggested that they are small and substantially weaker than those of *g*. A recent meta-analysis, for example, reported a meta-analytic correlation of $\bar{r} = .10$ between growth mindset and academic achievement (Sisk, Burgoyne, Sun, Butler, & Macnamara, 2018). A separate meta-analysis of mindset intervention effects in Sisk et al. (2018) revealed similar results, with very small overall effects on academic achievement ($\bar{d} = 0.08$). That said, there was substantial heterogeneity, with larger effects for students from low-socioeconomic status (SES) households ($\bar{d} = 0.34$), and non-significant effects for the remainder of students ($\bar{d} = 0.03$). It would thus be essential to consider the target population when evaluating the potential impact of a mindset intervention.

Even so, it is important to note that any observed effects of these interventions may or may not be due to mindset per se. That is, the Sisk et al. (2018) meta-analysis revealed that mindset interventions that successfully increased students' growth mindset had no overall effect on their academic achievement. Rather, mindset interventions that *failed* to alter mindset (or did not measure whether they changed students' mindset) appeared to drive the observed effects on academic achievement (Sisk et al., 2018). This raises the possibility that the "active ingredient" in mindset interventions is not mindset, but perhaps the encouragement of hard work, persistence in school, or other motivational factors. Consistent with this possibility, prior research indicated that mindset, grit, and locus of control loaded on a common "self-determination" factor, which was distinct from a second factor reflecting cognitive ability (Burgoyne, Hambrick, Moser, & Burt, 2018; Malanchini, Engelhardt, Grotzinger, Harden, & Tucker-Drob, 2018; Tucker-Drob, Briley, Engelhardt, Mann, & Harden, 2016). Furthermore, a recent mindset intervention appeared to increase participants' internal locus of control and challenge-seeking behavior, but had no effect on cognitive ability test performance (Burgoyne et al., 2018).

1.1. How does the mindset intervention exert its effects?

If the mindset intervention does indeed consistently alter motivational predictors of achievement, the question of how it does so remains. These kinds of questions are typically addressed via efforts to uncover the "active ingredients" of an intervention. While such work is clearly important, as demonstrated above, it nevertheless provides only a partial answer to the question. What is still totally unknown (for either the growth mindset intervention or any other intervention) is whether and how the genetic and environmental etiology of the outcome is affected by the intervention.

Etiology refers to the underlying causes of a given outcome. In behavioral genetics research, etiology typically refers to the genetic and environmental contributions to individual differences in an observed trait (i.e., a phenotype). Twin studies are one approach researchers use to illuminate the etiology of a trait. Twin studies leverage the differing degrees of genetic similarity between identical (monozygotic; MZ) twins, who share 100% of their genes, and fraternal (dizygotic; DZ) twins, who share an average of 50% of their segregating genes, to estimate genetic and environmental contributions to individual differences in a phenotype. Using this approach, etiology can be described in terms of "ACE" estimates, reflecting the proportion of variance in a phenotype attributable to (A) additive genetic effects, (C) shared environmental effects, which include all non-genetic influences that make siblings similar, and (E) non-shared environmental effects, which include all non-genetic influences that make siblings different, plus measurement error (Knopik et al., 2017). Because MZ twins are identical genetically, differences among them are thought to reflect nonshared or person-specific environmental influences. Furthermore, genetic influences can be inferred from the extent to which MZ twins are more similar than DZ twins on a given trait.

Importantly, ACE estimates are specific to a particular population at a particular time and can differ across real-world environments (Knopik et al., 2017; Turkheimer, Haley, Waldron, d'Onofrio, & Gottesman, 2003). As a case in point, Turkheimer et al. (2003) found that the genetic contribution to differences in children's intelligence was larger in higher-SES households than lower-SES households, while the contribution of the shared environment was smaller. Turkheimer et al.'s (2003) results are an example of a genotype-environment ($G \times E$) interaction, because the effect of the genotype on the phenotype differed depending on the environment.

[1] These estimates are typical for medium-to-high SES samples; however, as we discuss below, Turkheimer et al. (2003) found much lower heritability estimates for low SES samples.

How might an intervention alter the etiology of an outcome? One possibility is that the intervention increases genetic influences, perhaps by creating environmentally neutral conditions (in which everyone has the same recent exposure) or environmentally supportive conditions that allow genetic predispositions to have a more prominent effect. Alternatively, we could observe increases in shared environmental influences following the growth mindset intervention, that may or may not be accompanied by decreases in genetic influences. The latter findings would likely be interpreted as something akin to an environmental "direct effect," such that common exposure to the growth mindset intervention increases twin similarity regardless of their level of genetic similarity (for further discussion, see Burt, Plaisance, & Hambrick, 2019). Regardless, such work goes far beyond both the presence of mean changes in the outcome and the identification of the active ingredients supporting that change to examine foundational shifts in the genetic and environmental architecture of the outcome. This approach has practical implications for prevention and intervention science and also stands to illuminate the potential etiologic range of outcomes. To sum up, when interventions work as intended, are they altering the etiology of the outcome? And if so, how?

The new field of experimental behavior genetics (Burt et al., 2019) was proposed to answer exactly these kinds of questions, linking experimental science with traditional twin designs to answer questions about intervention effects on etiology. Such work builds on prior research in several ways, not the least of which is that it addresses a seemingly paradoxical finding from behavioral genetics research — even traits that are highly heritable are responsive to environmental intervention (Dickens & Flynn, 2001). For example, although academic achievement is highly heritable, randomized experiments have shown that it is influenced by class size (Nye, Hedges, & Konstantopoulos, 2000), the proportion of female students in the classroom (Whitmore, 2005), and school quality (Hastings & Weinstein, 2008). The seemingly counterintuitive nature of these findings hinges on a few implicit (but likely misguided) assumptions: (1) heritability runs counter to malleability, such that environmental interventions must be protracted and/or particularly strong to alter genetic effects, and (2) results from experimental science and correlational science do not inform each other (see Sauce & Matzel, 2013, for an excellent review). With respect to the latter, behavioral genetic work that points to the presence of genetic effects on individual differences in achievement is viewed as simply unrelated to experimental work pointing to mean increases in achievement with reduced class size, reflecting the distinct statistical foci of correlational versus experimental research. Namely, the effects of interventions are typically assessed via changes in the mean, whereas twin studies rely on decomposition of variance.

We would argue that these assumptions should be considered hypotheses to be rigorously tested (see, e.g., Lewontin, 1974; Longino, 2013; Tabery, 2014), and that efforts should be made to reconcile the respective findings of experimental and correlational science, thereby advancing not only our understanding of interventions, but also our understanding of etiology. Such work would move beyond analyses of "what is" — the *de novo* etiology of academic achievement-related constructs as they exist in nature — to shed much-needed light on "what could be" — how these *de novo* etiologies might change in response to environmental intervention (Burt et al., 2019).

The present study will do just this for the first time, evaluating the results of a randomized intervention embedded in a behavioral genetic design, as previously discussed in Burt et al. (2019). We specifically addressed two primary research questions: (1) What effect does a brief, online mindset intervention have on participants' mindset, along with grit, locus of control, challenge-seeking behavior, and cognitive ability test performance? (2) Does the etiology of mindset and other academic achievement-related constructs change due to a brief, online intervention? To address these questions, we conducted a short-term randomized control trial by administering a mindset intervention or active control materials to a large sample of twins that were oversampled for disadvantage. We measured participants' mindset, grit, locus of control, challenge-seeking behavior, and cognitive ability before and after the intervention. We then used multilevel modeling to investigate the behavioral effects of the mindset intervention and a series of twin models to assess the extent to which the etiologies of the achievement-related constructs shifted as a result of the intervention.

2. Method

2.1. Participants

The study was approved by the Michigan State University Institutional Review Board (#X17-302e, entitled "Understanding 'What Could Be': Heritability in a Randomized Twin Design"). Children provided informed assent, and parents provided informed consent for themselves and their children. Participants included 1668 twins from 840 twin families that were recruited from the Michigan State University Twin Registry (MSUTR), a large-scale, population-based registry of child, adolescent, and adult twins and their families. Because prior meta-analytic work (Sisk et al., 2018) indicated that mindset interventions exert larger effects on youth from disadvantaged scholastic and achievement backgrounds, we oversampled for neighborhood disadvantage. Participants ranged in age from 8 to 19 years ($M = 13.45$, $SD = 1.83$; over 99% were between the ages of 11 and 19) at the time of the study and 47.3% identified as female. Most participants (86.6%) identified as White, 5.5% identified as multiracial, 4.1% identified as Black, 0.8% identified as Asian, and 0.4% identified as American Indian; the majority of participants (98.5%) identified as non-Hispanic. Parental educational attainment ranged from completion of 8th grade or less to completion of a graduate degree.

Twin families were recruited from one of two studies. Specifically, 377 families (44.88% of the total sample) were recruited from another study within the MSUTR, the Twin Study of Behavioral and Emotional Development in Children (TBED-C). The TBED-C includes a population-based sample (mean annual household income: $79,602; median: $75,000) and an independent at-risk sample for which inclusion criteria specified that participating twin families lived in neighborhoods with modest to severe poverty (mean income: $57,359; median: $55,000; Burt & Klump, 2013). By comparison, the median annual income for families in Michigan is $75,703, and the living wage for Michigan families with two children is $56,186, according to 2019 Census data. To be eligible for participation in the TBED-C, neither twin could have a cognitive or physical condition that would preclude completion of the assessment (as

determined via parental screen; e.g., a significant developmental delay). The mean income in these 377 TBED-C families was $67,911 (median: $65,000; range: $0 to $250,000).

The remaining 463 families (55.12% of the sample) were recruited from the Michigan Twins Project (MTP), a recruitment registry of more than 32,000 twins embedded in the MSUTR and recruited through birth records. Across both samples, mean annual household income was $93,545 (median: $90,000), ranging from 0 to $300,000. Participants in the TBED-C and MTP were recruited through the mail for the current study. We collaborated with the Department of Vital Records in the Michigan Department of Health and Human Services to recruit twins in our specified age range. Pre-made recruitment packets were then mailed on our behalf by the Michigan Department of Health and Human Services to parents. A reply postcard was included for parents to indicate their interest in participating. Interested families were contacted directly by project staff. Parents who did not respond to the first mailing were sent additional mailings approximately one month apart until either a reply was received or up to four letters had been mailed. This recruitment strategy yielded an overall response rate of 58% for families across both samples, which is similar to or better than that of population-based twin registries that use anonymous recruitment mailings (Baker, Barton, & Raine, 2002; Hay, McStephen, Levy, & Pearsall-Jones, 2002).

2.2. Zygosity determination

Parents completed a standard zygosity questionnaire on behalf of the twins which included questions about physical similarity (Peeters, Van Gestel, Vlietinck, Derom, & Derom, 1998). On average, physical similarity questionnaires have accuracy rates of at least 95% when compared to DNA. The current study included 176 MZ male pairs (20.95% of the sample), 179 MZ female pairs (21.31% of the sample), 260 DZ male pairs (30.95% of the sample), 213 DZ female pairs (25.36% of the sample), and seven males and five females whose co-twins did not participate.

2.3. Measures

2.3.1. Demographic questionnaire

Participants' age, gender, race, and ethnicity were obtained from demographic questionnaires completed by participants in the MTP and TBED-C.

2.3.2. Mindset

This 3-item questionnaire assessed whether participants believe that their intelligence is fixed or malleable (Yeager et al., 2016). Participants responded to items such as "You have a certain amount of intelligence and you really can't do much to change it" using a 6-point Likert scale, with response options ranging from "Strongly disagree" to "Strongly agree." Higher scores on this measure correspond to more growth mindset.

2.3.3. Grit

This 8-item questionnaire assessed trait-level consistency of interests and perseverance of effort (Duckworth & Quinn, 2009). Participants responded to items such as "I often set a goal but later choose to pursue a different one" using a 5-point Likert scale, with response options ranging from "Very much like me" to "Not like me at all." Higher scores on this measure correspond to more grit.

2.3.4. Locus of control

This 28-item questionnaire assessed the extent to which participants believe that their academic performance is a result of internal or external factors (Trice, 1985). Participants reported whether items such as "College grades most often reflect the effort you put into classes," or "I have taken a course because it was an easy good grade at least once" are true or false as they relate to themselves. Higher scores on this measure correspond to more internal attributions.

2.3.5. Make-a-math worksheet

Participants were asked to construct a math worksheet by selecting a set of math problems from 3 different content areas. Participants were provided with the following description of the task:

> *What kind of math worksheet would you prefer? We are interested in what kinds of problems 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 session you will have the opportunity to answer these math problems. 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.*

The math problems were labeled with the descriptors "Very challenging problem," "Somewhat challenging problem," and "Not very challenging problem." An overall challenge-seeking score, intended to reflect challenge-approach motivation, was calculated for each participant by subtracting the number of easy problems selected for the worksheet from the number of very challenging problems selected for the worksheet. After completing the Make-A-Math Worksheet, participants were provided with the following statement:

Thank you for selecting the problems. Unfortunately, there is not enough time for you to complete the problems that you selected. However it is very helpful to know what kinds of problems you would have liked to work on, if there had been enough time. Thank you for your responses.

2.3.6. Cognitive ability

Two tests were administered to participants to measure crystallized and fluid cognitive ability or intelligence, respectively. The Shipley-2 Vocabulary test consisted of 40 items (Shipley, Gruber, Martin, & Klein, 2009). Participants attempted to identify the word most similar in meaning to the target word out of 4 options. At pre-test, participants were presented with the 20 odd-numbered items; at post-test, participants were presented with the 20 even-numbered items. The Shipley-2 Block Design test consisted of 26 items (Shipley et al., 2009). Participants attempted to identify the missing section(s) of an abstract design from a set of options, such that the completed design matched the target design. The same 26 items were presented to participants at pre-test and at post-test. No accuracy feedback was provided. We chose to use vocabulary and block design as measures of cognitive ability because pilot testing (Burgoyne et al., 2018) indicated that they could be administered in the same online survey platform as the intervention.

2.4. Composite variables

A self-determination composite variable was formed by averaging standardized scores on mindset, grit, locus of control, and challenge-seeking behavior. A cognitive ability composite variable was formed by averaging standardized scores on the vocabulary and block design tests.

2.5. Mindset intervention

The mindset intervention was adapted from Yeager et al. (2016) and included two conditions: a growth mindset condition and an active control condition. In both conditions, participants were presented with a lay-person "scientific review" article complete with graphics, compelling stories (e.g., Phineas Gage), and celebrity quotes. Participants were asked to read the entire article and remember the main points for a memory test. After reading the article, participants were asked to write a summary of the article and to rate the extent to which the article was difficult to read, credible, and persuasive, and how much they agreed with the article's points. The intervention content took approximately 25 min to complete.

The two conditions differed in terms of the content presented to participants. In the growth mindset condition, participants were presented with content suggesting that intelligence is developed from stimulating environments and can be improved with hard work (e.g., "the brain is like a muscle—it gets stronger (and smarter) when you exercise it") (Fig. 1). In the control condition, participants

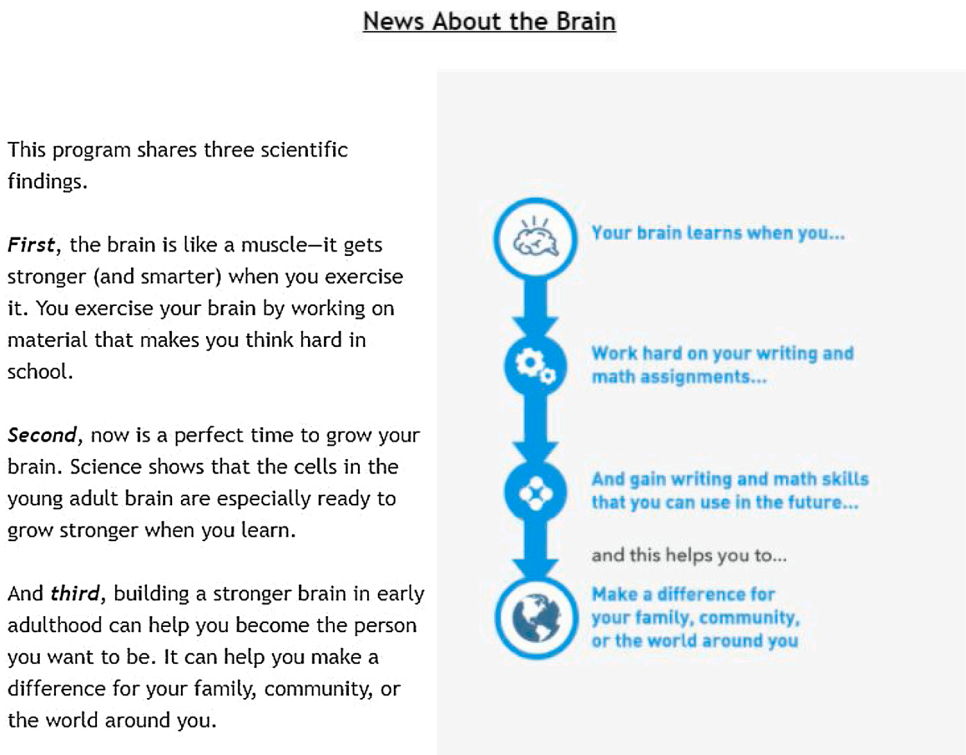

Fig. 1. Example materials from the growth mindset intervention.

were presented with content that reviewed basic findings about the human brain, but did not include statements indicating that intelligence is malleable (e.g., "the parietal lobe is where the brain interprets the sense of touch") (Fig. 2).

2.6. Procedure

Twins were randomly assigned to either the growth mindset intervention condition ($n = 818$) or the active control condition ($n = 850$) using a tracking table. Thus, for a given twin pair, there were three possible condition assignments: both twins received the mindset intervention ($n = 262$ pairs; 31.64% of the sample; $M_{age} = 13.26$, $SD = 1.80$; 50.4% female; 89% White, 4.6% Multiracial, 3.4% Black, 1% Asian, 0% American Indian; 99% non-Hispanic), both twins received the active control materials ($n = 279$ pairs; 33.70% of the sample; $M_{age} = 13.45$, $SD = 1.68$; 47% female; 84.2% White, 6.5% Multiracial, 4.7% Black, 1% Asian, 0.4% American Indian; 98.2% non-Hispanic), or one twin received the mindset intervention while the co-twin received the active control materials ($n = 287$ pairs; 34.66% of the sample; $M_{age} = 13.61$, $SD = 1.97$; 44.9% female; 87.1% White, 5.6% Multiracial, 3.8% Black, 0.3% Asian, 0.7% American Indian; 98.6% non-Hispanic). Participants were told that the purpose of the study was to examine relationships among personality, beliefs, emotions, and cognitive ability.

The experiment was conducted online via Qualtrics in a single session with three phases. In the first phase, participants completed the following pre-test measures, listed in order of administration: Vocabulary, Block Design, Mindset, Grit, Locus of Control, and the Make-A-Math Worksheet. In the second phase, participants were given either the growth mindset intervention or the active control materials. In the third phase, participants completed the following post-test measures: Vocabulary, Block Design, Mindset, Grit, Locus of Control, and the Make-A-Math Worksheet. Following the third phase, all participants were debriefed. Participants in the active control condition were provided with the growth mindset intervention materials as part of the debriefing process to ensure that all participants could benefit from the possible advantages of the intervention.

2.7. Analyses

We used multilevel modeling to assess the phenotypic main effects of time (pre-intervention and post-intervention) and experimental condition (growth mindset intervention and active control), as well as the interaction between time and experimental condition, on each of the outcome measures. In these models, time was nested in participant, and time, condition, and time × condition were treated as fixed coefficients. Age, gender, and race were treated as fixed covariates in the models. Our research question regarding mean changes in the outcome is most directly tested through the time × condition interaction, which indicates whether the scores changed significantly from baseline to follow-up and whether any observed changes were greater for participants completing the

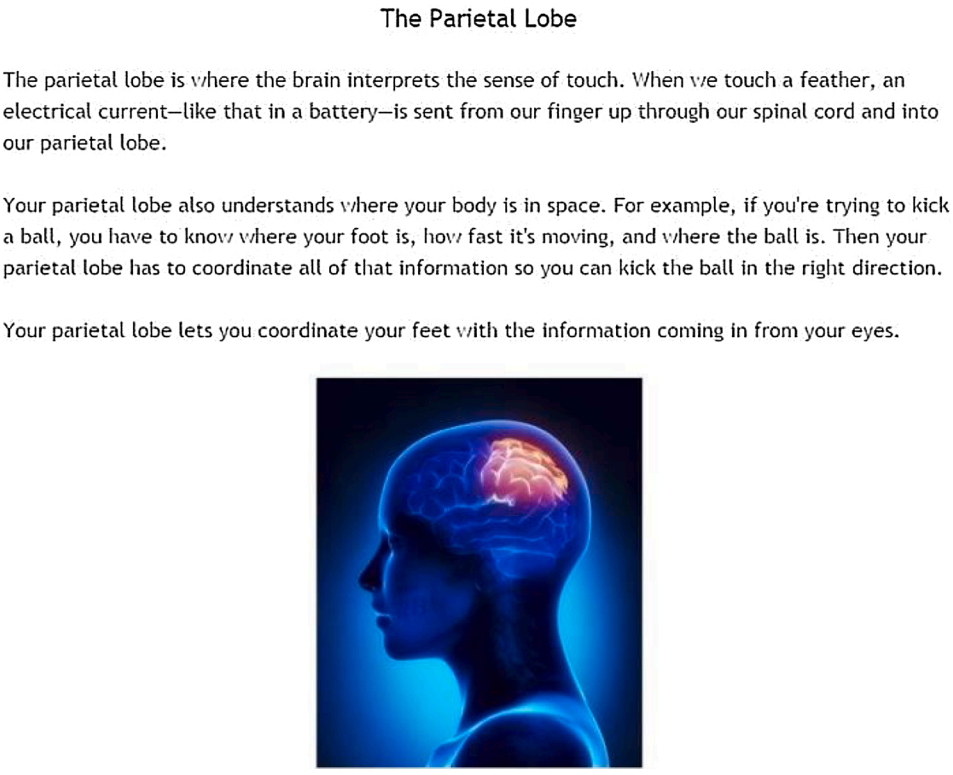

Fig. 2. Example materials from the active control condition.

intervention than for control participants. In addition, we conducted within-subjects *t*-tests, computing standardized effect sizes using the following formula: Cohen's $d = (M_{\text{post-test}} - M_{\text{pre-test}})/SD_{\text{pre-test}}$. As a supplemental analysis, we also examined whether the effects of the intervention were moderated by SES, as measured by the percentage of families residing below the poverty line according to 2017 US Census data, with the prediction that stronger effects would emerge for lower-SES participants.

For the behavioral genetic analyses, we leveraged the differing degrees of genetic similarity between MZ twins and DZ twins to determine the genetic and environmental contributions to individual differences in each of the study measures. These analyses, using the univariate classical twin model in M*plus* 8.0 (Muthén & Muthén, 2019), yielded estimates of (A) additive genetic effects, (C) shared environmental effects, and (E) nonshared environmental effects (including measurement error) on scores on each measure prior to the intervention.

Next, we used Mx, a structural equation modeling program (Neale, Boker, Xie, & Maes, 2003), to evaluate the effect of the intervention on etiology. We fit a univariate G × E twin model (Purcell, 2002) to measures for which the intervention had a phenotypic effect (i.e., altered the mean of a measure differentially for the intervention and control groups). Condition (growth mindset intervention vs. active control) was entered as a moderator, coded so that 0 = control and 1 = intervention. In these analyses, twins are not required to be concordant on the value of the moderator because treatment status was randomly distributed. The first, and least restrictive, model allows for simultaneous linear moderation of the A, C, and E parameters. We subsequently tested more restrictive models, constraining the linear moderators to zero, and evaluating the changes in model fit.

We determined the best-fitting model via a series of fit statistics. The baseline index of fit (minus twice the log-likelihood; −2lnL) was computed by estimating the means, variances, and covariances of the raw data. Model fit for the biometric G × E models was then evaluated with four indices: the Akaike information criterion (AIC; Akaike, 1987), Bayesian information criterion (BIC; Raftery, 1995), sample-size adjusted Bayesian information criterion (SABIC; Sclove, 1987), and Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). For all indices, lower values indicate better model fit. The best-fitting model was indicated by the lowest or most negative values for at least three of the four fit indices. Consistent with prior recommendations (Purcell, 2002), all reported parameter estimates from the G × E models are unstandardized so as not to obscure absolute changes with the moderator. To facilitate interpretation of these unstandardized estimates, outcome measures were standardized to a mean of zero and standard deviation of one prior to analyses. In addition, age, gender, and race were regressed out of the data for all behavioral genetic analyses, consistent with prior recommendations (McGue & Bouchard, 1984). As supplemental analyses, we also examined whether the effect of the intervention on etiology was further moderated by SES via a two-moderator extension of the univariate G × E model.

3. Results

3.1. Descriptive statistics

Correlations between measures at baseline are shown in Table 1. A growth mindset was associated with higher levels of grit ($r = .21$, $p < .001$) and more internal locus of control ($r = .25$, $p < .001$). The correlation between growth mindset and challenge-seeking behavior as measured by the make-a-math worksheet was also significant, albeit small ($r = .11$, $p < .001$). Furthermore, growth mindset was positively associated with scores on vocabulary ($r = .18$, $p < .001$) and weakly associated with scores on block design ($r = .07$, $p = .003$). Challenge-seeking behavior was not significantly associated with scores on vocabulary ($r = .04$, $p = .085$) or block design ($r = .04$, $p = .140$).

As expected, age was positively correlated with pre- and post-intervention scores on vocabulary ($F(1, 815.29) = 170.95$, $p < .001$, $r = .33$), block design ($F(1, 817.91) = 12.09$, $p < .01$, $r = .09$), and the cognitive ability composite variable ($F(1, 821.01) = 110.86$, $p < .001$, $r = .28$). Age was also positively correlated with the number of easy items chosen for the make-a-math worksheet ($F(1, 814.45) = 147.11$, $p < .001$, $r = .32$) and number of hard items chosen ($F(1, 810.08) = 169.07$, $p < .001$, $r = .35$). By contrast, age correlated negatively with locus of control ($F(1, 815.65) = 33.48$, $p < .001$, $r = -.18$); older participants reported more external locus of control. Female participants reported more internal locus of control ($F(1, 814.34) = 8.79$, $p < .01$, $d = 0.22$).

Table 1
Correlations at Time 1 ($N = 1668$).

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Mindset	–								
2. Grit	**.21**	–							
3. Locus of Control	**.25**	**.57**	–						
4. Easy Items	−.03	**−.07**	**−.15**	–					
5. Hard Items	**.12**	**.17**	**.19**	**.06**	–				
6. Challenge Seeking	**.11**	**.17**	**.25**	**−.71**	**.66**	–			
7. Self-Determination Composite	**.60**	**.73**	**.78**	**−.36**	**.43**	**.57**	–		
8. Vocabulary	**.18**	**.10**	**.08**	**.13**	**.20**	.04	**.12**	–	
9. Block Design	**.07**	.05	**.06**	**.10**	**.15**	.04	**.10**	**.21**	–
10. Cognitive Ability Composite	**.16**	**.08**	**.09**	**.13**	**.21**	**.06**	**.15**	**.78**	**.78**

Note. Bold, $p < .05$.

3.2. Phenotypic results

Mean scores at both time points and the test of the Time × Condition interaction are reported in Table 2. Mindset scores did not differ by condition at Time 1 ($F(1, 3175.80) = 1.66$, $p = .20$). Collapsing across condition, mindset scores increased significantly from Time 1 to Time 2 ($F(1, 2437.44) = 37.17$, $p < .001$, $d = 0.16$), indicating greater endorsement of growth mindset at the end of the study. More importantly, however, there was a significant interaction between time and condition ($F(1, 2437.44) = 15.11$, $p < .001$), such that participants completing the growth mindset intervention reported a greater shift toward a growth mindset ($d = 0.25$) from Time 1 to Time 2 than did control participants ($d = 0.06$). Grit scores did not differ by condition at Time 1 ($F(1, 3180.49) = 0.28$, $p = .60$). Collapsing across condition, grit increased from Time 1 to Time 2 ($F(1, 2439.21) = 9.39$, $p < .01$, $d = 0.08$), but there was no interaction between time and condition ($F(1, 2439.21) = 0.35$, $p = .55$). Locus of control scores did not differ by condition at Time 1 ($F(1, 3251.12) = 0.18$, $p = .68$). Collapsing across condition, locus of control did not change from Time 1 to Time 2 ($F(1, 2437.38) = 2.90$, $p = .09$), and there was no interaction between time and condition ($F(1, 2437.38) = 0.99$, $p = .32$).

Prior to the intervention, control participants selected more easy items on the make-a-math worksheet than treatment participants ($F(1, 3179.00) = 9.31$, $p < .01$, $d = 0.06$). There was no main effect of time ($F(1, 2435.53) = 0.42$, $p = .52$), or interaction between time and condition ($F(1, 2435.53) = 1.52$, $p = .22$). The number of hard items participants chose did not differ across groups at Time 1 ($F(1, 3248.23) = 1.07$, $p = .30$). Collapsing across condition, the number of hard items participants chose increased from Time 1 to Time 2 ($F(1, 2430.93) = 20.07$, $p < .001$, $d = 0.10$), but there was no interaction between time and condition ($F(1, 2430.93) = 2.29$, $p = .13$). Overall challenge-seeking scores on the make-a-math worksheet (i.e., number of hard items – number of easy items) differed significantly at Time 1 across conditions ($F(1, 3166.01) = 7.17$, $p < .01$, $d = 0.01$), with participants in the treatment condition choosing more hard items relative to easy items than control participants. Collapsing across condition, scores increased significantly from Time 1 to Time 2 ($F(1, 2430.88) = 9.48$, $p < .01$, $d = 0.08$), indicating greater willingness to select challenging math problems among participants in both conditions. However, the interaction between time and condition was not significant ($F(1, 2430.88) = 3.03$, $p = .08$).

For the self-determination composite variable, there was no difference between conditions at Time 1 ($F(1, 3270.14) = 1.41$, $p = .24$). There was no main effect of time ($F(1, 2452.54) = 0.04$, $p = .84$), however, there was a significant interaction between time and condition ($F(1, 2452.54) = 8.24$, $p = .004$), such that participants in the treatment condition reported a slightly greater increase in self-determination from Time 1 to Time 2 ($d = 0.07$) than participants in the control condition ($d = -0.07$).

Vocabulary scores did not differ significantly by condition at Time 1 ($F(1, 3130.42) = 1.68$, $p = .20$). Collapsing across condition,

Table 2
Mean scores at Time 1 and Time 2.

Measure	Condition	Time 1 Mean (*SD*)	Time 2 Mean (*SD*)	Time × Condition Interaction
Mindset	Control Growth	3.42 (1.13) 3.38 (1.19)	3.49 (1.12) 3.68 (1.22)	$F(1, 2437.44) = 15.11$, $p < .001$
Grit	Control Growth	2.33 (0.62) 2.32 (0.59)	2.37 (0.67) 2.38 (0.64)	$F(1, 2439.21) = 0.35$, $p = .55$
Locus of Control	Control Growth	16.70 (4.70) 16.62 (4.68)	16.76 (4.72) 16.93 (4.66)	$F(1, 2437.38) = 0.99$, $p = .32$
Easy Items	Control Growth	7.24 (4.96) 6.93 (4.95)	7.33 (5.11) 6.71 (5.16)	$F(1, 2435.53) = 1.52$, $p = .22$
Hard Items	Control Growth	3.90 (4.82) 3.68 (4.45)	4.21 (5.01) 4.31 (5.05)	$F(1, 2430.93) = 2.29$, $p = .13$
Challenge Seeking	Control Growth	−3.34 (6.72) −3.25 (6.47)	−3.12 (7.00) −2.39 (7.13)	$F(1, 2430.88) = 3.03$, $p = .08$
Self-Determination	Control Growth	0.03 (2.67) −0.03 (2.66)	−0.16 (2.66) 0.16 (2.73)	$F(1, 2452.54) = 8.24$, $p = .004$
Vocabulary	Control Growth	13.49 (2.83) 13.18 (2.80)	12.65 (3.00) 12.35 (3.06)	$F(1, 2438.03) = 0.00$, $p = .97$
Block Design	Control Growth	10.80 (2.27) 10.54 (2.39)	11.12 (2.17) 10.96 (2.26)	$F(1, 2440.98) = 0.36$, $p = .55$
Cognitive Ability	Control Growth	0.11 (1.56) −0.11 (1.55)	0.08 (1.56) −0.09 (1.59)	$F(1, 2453.88) = 0.37$, $p = .54$

Note. Intervention group Time 1: $N = 818$, Time 2: $Ns = 814–818$; control group Time 1: $N = 850$; Time 2: $Ns = 849–850$. The Time × Condition interaction indicates whether the intervention altered responses relative to baseline, and did so differentially for those assigned to the intervention condition versus the control condition.

vocabulary scores decreased from Time 1 to Time 2 ($F(1, 2438.03) = 121.57$, $p < .001$, $d = -0.30$), but there was no interaction between time and condition ($F(1, 2438.03) = 0.00$, $p = .97$). Block design scores differed significantly by condition at Time 1 ($F(1, 2837.89) = 5.11$, $p < .05$, $d = 0.11$), with control participants scoring slightly higher than those in the treatment condition. Collapsing across condition, block design scores increased from Time 1 to Time 2 ($F(1, 2440.98) = 38.61$, $p < .001$, $d = 0.16$), but there was no interaction between time and condition ($F(1, 2440.98) = 0.36$, $p = .55$). For the cognitive ability composite variable, scores differed significantly by condition at Time 1 ($F(1, 3171.97) = 5.26$, $p < .05$, $d = 0.14$), with control participants scoring slightly higher. There was no main effect of time ($F(1, 2453.88) = 0.20$, $p = .66$) or interaction between time and condition ($F(1, 2453.88) = 0.37$, $p = .54$).

In sum, the intervention significantly altered participants' mindset and scores on the self-determination composite variable, but had no effect on grit, locus of control, challenge-seeking behavior, or performance on the cognitive ability tests. Supplemental analyses indicated that SES did not significantly interact with any of the mindset intervention's effects. We also examined the intervention's effects after excluding twin pairs who were discordant for condition (to reduce the likelihood of cross-condition communication about the intervention content), however, the same pattern of results emerged as in the full sample.

3.3. Twin model results

First, we used a simple univariate model to estimate genetic (A), shared environmental (C), and nonshared environmental (E) contributions to individual differences in each measure at Time 1. Results are shown in Table 3. There were significant genetic contributions to grit, locus of control, the number of easy items selected for the make-a-math worksheet, challenge-seeking scores on the make-a-math worksheet, the self-determination composite variable, and block design. For vocabulary scores, the variance in scores was unexpectedly explained entirely by environmental factors. Given that the heritability of verbal intelligence increases with age (Van Soelen et al., 2011), we conducted a follow-up analysis on adolescents age 14 and older to determine whether significant genetic effects on vocabulary scores would emerge in this subsample: they did not. Shared environmental factors contributed significantly to variance in the number of hard items chosen for the make-a-math worksheet, vocabulary, and the cognitive ability composite. Nonshared environmental factors were significant for all measures, which was expected given that they contain measurement error.

3.3.1. Intraclass correlations

As a preliminary test of etiologic moderation, we compared twin intraclass correlations separately by zygosity and treatment condition. We focused on mindset and the self-determination composite variable, because these were the only outcome measures that changed as a result of the intervention. The correlations are shown in Table 4 ($N = 541$ twin pairs; these analyses necessarily exclude twin pairs that were discordant for condition, although subsequent G × E analyses include all participating twin pairs).

For mindset, MZ and DZ twin similarity in the intervention group did not differ significantly at Time 1 ($z = 0.42$, $p = .67$) or Time 2 ($z = 1.44$, $p = .15$), but there was a trend such that MZ similarity increased ($z = 0.41$, $p = .68$) and DZ similarity decreased ($z = -0.63$, $p = .53$) across the two time points. Although these changes were non-significant, this pattern of results indicates a possible increase in additive genetic contributions to individual differences in mindset as a result of the intervention. In the control group, MZ and DZ similarity on mindset also did not differ significantly at Time 1 ($z = 1.67$, $p = .09$) or Time 2 ($z = 0.94$, $p = .35$), and MZ similarity decreased non-significantly across the study ($z = -0.67$, $p = .50$) while DZ similarity remained constant across the two time points ($z = 0.00$, $p = 1.00$).

For the self-determination composite variable, MZ and DZ twin similarity in the intervention group did not differ significantly at Time 1 ($z = 0.64$, $p = .52$) or Time 2 ($z = 1.18$, $p = .24$), and neither point estimate changed significantly across the two time points (MZ: $z = 0.18$, $p = .86$; DZ: $z = -0.38$, $p = .70$). In the control group, MZ and DZ twin similarity on the self-determination composite variable differed significantly before the intervention ($z = 3.76$, $p < .001$) and after the intervention ($z = 2.92$, $p < .01$). However, neither MZ nor DZ twin similarity changed significantly across the two time points ($z = -0.71$, $p = .48$ and $z = 0.09$, $p = .93$, respectively).

Table 3
ACE variance estimates at Time 1.

Measure	Standardized variance estimates		
	A	C	E
Mindset	.23†	.10	**.67**
Grit	**.35**	.00	**.65**
Locus of Control	**.27**	.15	**.58**
Easy Items	**.42**	.00	**.58**
Hard Items	.26†	**.22**	**.52**
Challenge Seeking	**.37**	.00	**.62**
Self-Determination Composite	**.47**	.00	**.53**
Vocabulary	.00	**.41**	**.59**
Block Design	**.32**	.00	**.68**
Cognitive Ability Composite	.23†	**.21**	**.57**

Note. $N = 824$ pairs. Bold, $p < .05$. †, $p < .10$.

Table 4
Twin intraclass correlations for control and intervention groups.

Measure	Time	Control		Intervention	
		*r*MZ	*r*DZ	*r*MZ	*r*DZ
Mindset	Time 1	.36	.17	.30	.25
	Time 2	.28	.17	.35	.18
Self-Determination	Time 1	.54	.14	.41	.34
	Time 2	.47	.15	.43	.30

3.3.2. $G \times E$ model results

Next, we conducted formal tests of etiologic moderation (Purcell, 2002), evaluating whether genetic, shared environmental, and/or nonshared environmental contributions to individual differences in mindset or the self-determination composite variable shifted as a consequence of intervention condition. If a model fit significantly worse when additive genetic moderation was constrained to zero, for example, this would indicate moderation of the additive genetic component of variance. Fit statistics and parameter estimates for etiologic moderation of mindset and self-determination are shown in Tables 5 and 6.

Prior to the intervention, the no-moderation model provided the best fit to the data for mindset according to all four indices. These findings suggest that the etiology of mindset did not vary by treatment condition at Time 1. In sharp contrast, after the intervention, the A-only moderation model best fit the data (results from this model are depicted in Fig. 3). These findings suggest that, relative to the control condition, the growth mindset intervention served to increase the additive genetic contribution to individual differences in mindset (moderator = .18, $p < .05$). Neither shared nor nonshared environmental contributions shifted as a result of the intervention, as both moderators could be constrained to zero without loss in model fit. These observed differences in genetic influences following the intervention, along with the fact that these differences were not present before the intervention, suggest that the etiologic differences across conditions emerged as a consequence of the intervention.

Next, we conducted supplemental analyses to evaluate whether the etiologic shift in mindset following the intervention varied with level of youth disadvantage (operationalized here via neighborhood poverty, or the percentage of families residing below the poverty line according to 2017 US Census data). Results argued against the presence of interactions between the etiologic moderators for the intervention and those for disadvantage. Fixing the joint ACE moderators to zero resulted in a non-significant chi-square change of 2.74 ($df = 3$, $p = .43$), indicating that the moderating effects of the intervention on the etiology of mindset did not vary with the level of youth disadvantage.

By contrast, the intervention did not appear to alter the etiology of the self-determination composite variable. At Time 1, the no-moderation model best fit the data, according to all fit indices, meaning that etiology did not differ across groups before the intervention. At Time 2, the no-moderation model still provided the best fit according to all fit indices, indicating that the etiology did not change as a result of the intervention. For example, in the A-only moderation model at Time 2, the moderator effect was −.04 ($p = .45$), indicating that there was no moderation of the additive genetic component for the self-determination composite variable.

3.3.3. Mechanisms of change

Given these results, we conducted a series of post-hoc analyses to investigate how the growth mindset intervention acted to increase genetic influences on mindset. We first evaluated whether the genetic variance in mindset at Time 1 overlapped with the genetic variance in mindset at Time 2 via a simple bivariate twin model. When used to examine longitudinal data, the bivariate model decomposes the covariance between two phenotypes to uncover the sources of etiologic stability and change over time. This first set of analyses focused on twin pairs concordant for condition. Regardless of whether twins were in the intervention group or the control group, we found no evidence of unique genetic influences at Time 2 (estimates were .00; both $ps = 1$). These results were echoed when

Table 5
Mindset and self-determination model fit statistics.

Time	Model	−2lnL	*df*	AIC	BIC	SABIC	DIC
Variable: Mindset							
Time 1	**No moderation**	**4530.38**	**1616**	**1298.38**	**−3147.01**	**−581.14**	**−1662.01**
	Linear ACE moderation	4527.81	1613	1301.81	−3138.25	−577.14	−1656.00
Time 2	No moderation	4531.49	1615	1301.49	−3143.11	−578.82	−1659.02
	Linear ACE moderation	4525.78	1612	1301.78	−3135.91	−576.39	−1654.58
	Linear A moderation only	**4525.94**	**1614**	**1297.94**	**−3142.53**	**−579.83**	**−1659.36**
Variable: Self-Determination							
Time 1	**No moderation**	**4514.58**	**1627**	**1260.58**	**−3196.75**	**−613.39**	**−1701.64**
	Linear ACE moderation	4512.89	1624	1264.89	−3187.54	−608.95	−1695.19
Time 2	**No moderation**	**4511.59**	**1626**	**1259.59**	**−3194.90**	**−613.13**	**−1700.70**
	Linear ACE moderation	4510.31	1623	1264.31	−3185.48	−608.48	−1694.04

Note. The best-fitting model for a given set of analyses is highlighted in bold font, and is indicated by the lowest AIC, BIC, SABIC, and DIC values for at least 3 of the 4 fit indices.

Table 6
Unstandardized path and moderation parameter estimates for the full linear moderation and best-fitting moderation models for mindset and self-determination.

Time	Model	Paths			Linear Moderators		
		a_1	c_1	e_1	A	C	E
Variable: Mindset							
Time 1	**No moderation**	**.48**	**.31**	**.82**	–	–	–
	Linear ACE moderation	**.56**	.20	**.77**	−.18	.23	.07
	No moderation	**.43**	**.33**	**.83**	–	–	–
Time 2	Linear ACE moderation	**.37**	.33	**.82**	.14	.00	.03
	Linear A moderation only	**.36**	**.32**	**.83**	**.18**	–	–
Variable: Self-Determination							
Time 1	**No moderation**	**.69**	.00	**.73**	–	–	–
	Linear ACE moderation	.40	**.68**	**.77**	.15	−.28	−.03
Time 2	**No moderation**	**.64**	.19	**.74**	–	–	–
	Linear ACE moderation	**.53**	.55	**.76**	.05	−.19	−.01

Note. Bold, $p < .05$.

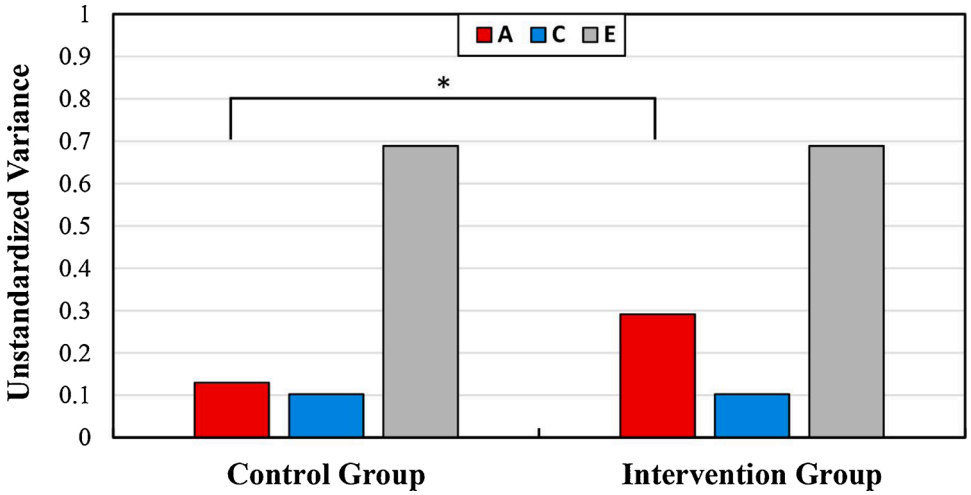

Fig. 3. The etiology of mindset at Time 2 for the control group and intervention group. The A parameter differed significantly across groups after the intervention, $p < .05$. A = additive genetic variance, C = shared environmental variance, E = nonshared environmental variance.

fitting the bivariate G × E model as well, which is able to further disambiguate the moderation terms across genetic variance unique to Time 2 and genetic variance that overlaps across Times 1 and 2. We found that the additive genetic moderator unique to Time 2 could be constrained to zero without a decrement in model fit, while the additive genetic moderator of variance common to Time 1 and Time 2 could not, further indicating the intervention acted to increase the importance of genetic influences that were already present. Put another way, because there were no novel genetic influences at Time 2, the intervention appears to have augmented preexisting genetic variance, rather than activating novel genetic influences that were not present prior to the intervention.

4. Discussion

In the present study, we administered a brief, online mindset intervention to a large sample of twins ($N = 1668$; $M_{age} = 13.5$). Participants were randomly assigned to either a mindset intervention condition or an active control condition. We measured participants' growth mindset, grit, locus of control, challenge-seeking behavior, and cognitive ability test performance before and after the intervention. We found that the mindset intervention successfully induced a growth mindset, but had no effect on grit, locus of control, challenge-seeking behavior, or cognitive ability test performance. Although the intervention also altered scores on a self-determination composite variable reflecting mindset, grit, locus of control, and challenge-seeking behavior, this was likely driven by the mindset intervention's effects on mindset—the trained trait—because the intervention did not significantly alter the means of any of the other motivational traits. These phenotypic results were not moderated by SES. Thus, the effects of the intervention were narrower than observed in our pilot study (Burgoyne et al., 2018), which used the same intervention and found effects on mindset, locus of control, and challenge-seeking in an older and more disadvantaged sample of 488 MTurk participants ($M_{age} = 21.9$).

We also estimated pre-intervention genetic and environmental contributions to individual differences in mindset, grit, locus of

control, challenge-seeking behavior, a self-determination composite variable, and cognitive ability, by comparing the similarity of MZ and DZ twins. We found that mindset was subject to small-to-moderate (albeit non-significant) genetic influences and moderate nonshared environmental influences, while shared environmental effects were negligible and non-significant. These findings are consistent with the results of Tucker-Drob et al. (2016), who found small genetic effects and large nonshared environmental effects on individual differences in mindset. For the other measures, there were small-to-moderate genetic contributions to everything except vocabulary scores, results which differ rather dramatically from those observed in other studies (Deary, Johnson, & Houlihan, 2009). Although we can only speculate as to why this was the case, it is possible that the vocabulary test was too brief (20 items per administration) to yield robust heritability estimates. Shared environmental factors contributed significantly to variance in the number of hard items chosen for the make-a-math worksheet, vocabulary scores, and the cognitive ability composite. As expected, nonshared environmental factors were significant for all measures.

Our final set of analyses assessed whether the mindset intervention altered the etiology of mindset or a self-determination composite variable using a series of univariate G × E analyses. Although the etiology of the self-determination composite was not moderated by the intervention, we did observe small but significant etiologic moderation of mindset as a result of the intervention. Specifically, a model in which additive genetic effects were moderated by condition at Time 2 fit better than a model that did not include moderation terms. The findings pointed to an increase in additive genetic effects for participants in the intervention condition at Time 2 compared to participants in the control condition at Time 2. This suggests, perhaps counterintuitively, that a randomly-assigned environmental intervention led to an increase in *genetic contributions* to mindset. Follow-up analyses further revealed that the genetic contributions to variance in mindset at Time 2 were not unique to Time 2, but rather were shared with the genetic contributions to variance in mindset at Time 1. Accordingly, the data collectively indicate that the intervention increased the contribution of pre-existing genetic influences on mindset from Time 1 to Time 2, rather than recruiting novel genetic influences at Time 2.

5. Implications

To our knowledge, this is the very first study to embed a randomized intervention trial within a traditional twin study design. We found that a brief mindset intervention not only successfully induced greater growth mindset, but also led to a foundational change in its etiology. The results call into question a handful of implicit assumptions underpinning most behavioral genetics research. In particular, our results show that environmental interventions need not be protracted or particularly strong to alter genetic effects. In other words, this study provides empirical evidence against the (often implicit) assumption that high heritability estimates indicate a lack of malleability. Second, this study shows that some heritability estimates appear to be considerably more malleable than is typically assumed, even following brief exposure to educational materials. Thus, rather than reflecting the genetic 'skeleton' of a given phenotype, heritability estimates are better viewed as snapshots of genetic influences on phenotypic variation at particular points in time, under particular environmental conditions, and with respect to particular populations.

Philosophers of science have advocated for precisely this way of understanding heritability, going back to Lewontin's (1974) paper on analysis of variance and analysis of causes (e.g., Plaisance, 2006; Tabery, 2014). While much of this philosophical work has argued that heritability estimates are not particularly stable, especially in new environments, our study is the first to actually *demonstrate* just how easily heritability estimates can sometimes be altered through short environmental interventions. Indeed, our new methodological approach is uniquely capable of such an empirical demonstration, even going beyond what Turkheimer et al. (2003) found in their study of IQ and SES, as a mindset intervention is a far more subtle environmental exposure than is SES.

This more nuanced understanding of genetic influences has clear implications for not only our interpretation of heritability estimates from twin studies, but also how we understand the results of genome-wide association studies (GWAS) vis-à-vis missing heritability (or the fact that the proportions of variance uncovered in GWAS are typically only a fraction of those identified in twin studies). Given how fickle heritability estimates can be, it is not quite so surprising that GWAS studies have not produced the outcomes expected by those who interpreted high heritability estimates in more traditional ways. Furthermore, the more nuanced understanding we present here applies to the other ACE components (shared and nonshared environmental variance) as well. Estimates of these components are also best seen as time-, environment-, and population-relative, as Turkheimer et al. (2003) demonstrate.

The current findings also strongly suggest that experimental science has the potential to meaningfully inform research in behavioral genetics. As noted, the field of behavioral genetics is focused almost exclusively on epidemiological evaluations of individual differences around the mean in naturally occurring environments, with very little attention paid to either the mean itself or to the effects of interventions. Our results suggest that broadening our methodologic focus to include experimental science can have important downstream consequences for our understanding of etiology. By the same token, these results also suggest that a correlational discipline like behavioral genetics can inform experimental science. Indeed, our results add a new wrinkle for researchers interested in identifying the "active ingredients" of an intervention, by demonstrating that intervention effects can include changes in the etiology of an outcome in addition to changes in mean scores. Our approach thus not only has practical implications for prevention and intervention science, but also stands to illuminate the potential etiologic range of outcomes (Burt et al., 2019).

All that said, the question of *how* the intervention increased genetic influences on mindset remains. One possibility is that the genetic contribution to individual differences in mindset became stronger in an environment that provided motivational support. This interpretation is also consistent with the bioecological model (Bronfenbrenner & Ceci, 1994), which suggests that genetic influences can sometimes increase in importance under supportive and protective environmental conditions and decrease under less ideal conditions (Burt et al., 2019). That said, the bioecological model also predicts a concomitant decrease in environmental effects alongside the increase in genetic effects, which was not observed in the present study.

Yet another possible explanation for our findings is that the increase in genetic effects on individual differences in mindset after the

intervention reflects genetically-influenced processes related to something besides mindset. During the intervention, participants were told that the brain is like a muscle that can grow stronger and smarter with effort, effectively priming them to endorse a growth mindset. In other words, participants may have surmised the purpose of the intervention and, intentionally or otherwise, altered their responses to support it (Burgoyne et al., 2018). Consistent with this possibility, we note that Burgoyne et al. (2018) found that mindset loaded on a factor with other motivational constructs prior to intervention, but loaded more strongly on a factor with cognitive ability measures after the intervention. This evidence is circumstantially consistent with the possibility that the mindset measure may have captured variance shared with general intelligence following the intervention. Despite this, we would argue against the conclusion that the current results reflect genetically-influenced demand characteristics, primarily because the intervention accentuated genetic influences that contributed to mindset de novo. In other words, the increase in genetic influences cannot reflect intervention demand characteristics since those influences were already operational prior to the intervention.

6. Limitations

There are two key limitations to the current study that should be considered. First, we note that it is unclear how enduring mindset intervention effects are at either the phenotypic or etiologic levels, as most studies measure effects within four months of intervention administration (Sisk et al., 2018). Another potential limitation concerns the extent to which the observed changes in etiology for mindset are attributable to variability in measurement over time. That is, repeated measurement might yield slightly different ACE estimates for mindset, which, by chance, were moderated by condition following the intervention. To address this possibility, we conducted ancillary ACE moderation analyses on the measures which were not significantly affected by the intervention at the phenotypic level (results not shown). There was no evidence of etiologic moderation by condition for grit, locus of control, challenge-seeking behavior, vocabulary, or block design scores, suggesting that measures which were unaffected by the intervention at the phenotypic level were also unaffected by the intervention at the etiologic level. These findings provide disconfirmatory evidence to an interpretation of our results which holds that the etiologic moderation of mindset by the intervention was simply due to variation in measurement over time, and also serve to strengthen the connection between the etiologic and phenotypic effects of the intervention on mindset.

7. Conclusion

In addition to shedding light on the etiology of academic achievement-related constructs, the present study represents a first attempt to address the question “what could be?”, or how etiology might change in response to environmental intervention. An important goal for future work is to further disentangle malleability and heritability by conducting experimental behavioral genetic interventions with effects on multiple outcomes differing in de novo etiology, and testing for moderation by environmental factors such as disadvantage. As a field, experimental behavioral genetics (Burt et al., 2019) promises one path for reconciling the respective findings of experimental and correlational science, thereby advancing not only our understanding of interventions, but also broadening our understanding of etiology.

CRediT authorship contribution statement

Alexander P. Burgoyne: Conceptualization, Writing - original draft, Writing - review & editing, Visualization. **Sarah Carroll:** Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **D. Angus Clark:** Formal analysis, Writing - review & editing. **David Z. Hambrick:** Conceptualization, Writing - review & editing, Funding acquisition. **Kathryn S. Plaisance:** Conceptualization, Writing - review & editing, Funding acquisition. **Kelly L. Klump:** Writing - review & editing. **S. Alexandra Burt:** Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

This research was funded by the Genetics of Human Agency grant from the Templeton Foundation. We thank Eric Turkheimer for supporting this research.

References

Akaike, H. (1987). Factor analysis and AIC. *Selected papers of Hirotugu Akaike* (pp. 371–386). New York, NY: Springer.
Baker, L. A., Barton, M., & Raine, A. (2002). The southern California twin register at the University of Southern California. *Twin Research and Human Genetics, 5*, 456–459.
Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nuture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101*, 568–586.
Burgoyne, A. P., Hambrick, D. Z., Moser, J. S., & Burt, S. A. (2018). Analysis of a mindset intervention. *Journal of Research in Personality, 77*, 21–30.
Burgoyne, A. P., Hambrick, D. Z., & Macnamara, B. N. (2020). How firm are the foundations of mindset theory? The claims appear stronger than the evidence. *Psychological Science, 31*, 258–267.
Burt, S. A., & Klump, K. L. (2013). The Michigan state university twin registry (MSUTR): An update. *Twin Research and Human Genetics, 16*, 344–350.
Burt, S. A., Plaisance, K. S., & Hambrick, D. Z. (2019). Understanding “what could be”: A call for ‘Experimental Behavioral Genetics’. *Behavior Genetics, 49*, 235–243.
Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S. H., et al. (2012). Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behavior Genetics, 42*, 699–710.

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology, 113*, 492–511.
Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21.
Deary, I. J., Johnson, W., & Houlihan, L. M. (2009). Genetic foundations of human intelligence. *Human Genetics, 126*, 215–232.
Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108*, 346–369.
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., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*, 1087–1101.
Dweck, C. S. (2008). *Mindset: The new psychology of success*. Ballantine Books.
Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256–273.
Findley, M. J., & Cooper, H. M. (1983). Locus of control and academic achievement: A literature review. *Journal of Personality and Social Psychology, 44*, 419–427.
Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics, 123*, 1373–1414.
Hay, D. A., McStephen, M., Levy, F., & Pearsall-Jones, J. (2002). Recruitment and attrition in twin register studies of childhood behavior: The example of the Australian twin ADHD project. *Twin Research and Human Genetics, 5*, 324–328.
Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2017). *Behavioral genetics*. New York, NY: Worth Publishers.
Lewontin, R. C. (1974). Annotation: The analysis of variance and the analysis of causes. *American Journal of Human Genetics, 26*, 400–411.
Li, Y., & Bates, T. C. (2019). You can't change your basic ability, but you work at things, and that's how we get hard things done: Testing the role of growth mindset on response to setbacks, educational attainment, and cognitive ability. *Journal of Experimental Psychology: General, 148*, 1640–1655.
Longino, H. E. (2013). *Studying human behavior: How scientists investigate aggression and sexuality*. University of Chicago Press.
Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P., & Tucker-Drob, E. M. (2018). "Same but different": Associations between multiple aspects of self-regulation, cognition, and academic abilities. *Journal of Personality and Social Psychology, 117*, 1164–1188.
McGue, M., & Bouchard, T. J. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics, 14*(4), 325–343.
Muthén, L. K., & Muthén, B. (2019). *Mplus. The comprehensive modelling program for applied researchers: User's guide*, 5.
Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling*. Richmond, VA: Department of Psychiatry. Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University.
Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37*, 123–151.
Peeters, H., Van Gestel, S., Vlietinck, R., Derom, C., & Derom, R. (1998). Validation of a telephone zygosity questionnaire in twins of known zygosity. *Behavior Genetics, 28*, 159–163.
Plaisance, K. S. (2006). *Behavioral genetics and the environment: The generation and exportation of scientific claims*. University of Minnesota.
Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry, 20*, 98–108.
Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research and Human Genetics, 5*, 554–571.
Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.
Sauce, B., & Matzel, L. D. (2013). The causes of variation in learning and behavior: Why individual differences matter. *Frontiers in Psychology, 4*, 395.
Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333–343.
Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *Shipley-2 manual*. Los Angeles, CA: Western Psychological Services.
Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses. *Psychological Science, 29*, 549–571.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*, 583–639.
Tabery, J. (2014). *Beyond versus: The struggle to understand the interaction of nature and nurture*. MIT Press.
Trice, A. D. (1985). An academic locus of control scale for college students. *Perceptual and Motor Skills, 61*, 1043–1046.
Tucker-Drob, E. M., Briley, D. A., Engelhardt, L. E., Mann, F. D., & Harden, K. P. (2016). Genetically-mediated associations between measures of childhood character and academic achievement. *Journal of Personality and Social Psychology, 111*, 790–815.
Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*, 623–628.
Van Soelen, I. L., Brouwer, R. M., Van Leeuwen, M., Kahn, R. S., Pol, H. E. H., & Boomsma, D. I. (2011). Heritability of verbal and performance intelligence in a pediatric longitudinal sample. *Twin Research and Human Genetics, 14*, 119–128.
Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review, 95*, 199–203.
Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., et al. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of Educational Psychology, 108*, 374–391.