Large Studies Reveal How Reference Bias Limits Policy Applications of Self-Report Measures

Benjamin Lira[a,1,2], Joseph M. O'Brien[b,1], Pablo A. Peña[c], Brian M. Galla[d], Sidney D'Mello[e], David S. Yeager[b], Amy Defnet[f], Tim Kautz[f], Kate Munkacsy[f], and Angela L. Duckworth[a]

[a]University of Pennsylvania; [b]University of Texas at Austin; [c]University of Chicago; [d]University of Pittsburgh; [e]University of Colorado - Boulder; [f]Mathematica

This manuscript was compiled on May 25, 2022

There is growing policy interest in identifying contexts that cultivate self-regulation. Doing so often entails comparing groups of individuals (e.g., from different schools). We show that self-report questionnaires—the most prevalent modality for assessing self-regulation—are prone to *reference bias*, defined as systematic error arising from differences in the implicit standards by which individuals evaluate behavior. In three studies, adolescents (N = 229,685) whose peers performed better academically rated themselves lower in self-regulation and held higher standards for self-regulation. This effect was not observed for task measures of self-regulation and led to paradoxical predictions of college persistence six years later. These findings suggest that standards for self-regulation vary by social group, limiting the policy applications of self-report questionnaires.

reference bias | measurement | self-regulation | contextual effects | noncognitive skills

Self-regulation refers to a diverse set of personal qualities, distinct from cognitive ability, that enable individuals to set and pursue goals. A component of all major conceptual frameworks of personal qualities (1–7), self-regulation and its facets play a starring role in theories of child development (8–10) and adult personality (11) and psychopathology (12), as well as behavioral economics (13), and labor economics (14).The centrality of self-regulation across diverse traditions in behavioral science is justified by the discovery that individual differences in self-regulation predict later life outcomes, including academic performance (15–17); physical and mental health (18–20); well-being and life satisfaction (21); civic and social behavior (19, 22); job performance (23); earnings (19, 24–26); and wealth (19, 24). Moreover, the effects of self-regulation are independent of, and comparable in magnitude to, cognitive ability and family socioeconomic status (SES) (15, 19).

A half-century of basic research suggests that self-regulation develops optimally in caring environments that encourage adaptive goal-relevant knowledge (e.g., strategies for managing attention), beliefs (e.g., that emotion and motivation can be regulated), and values (e.g., that self-regulation is important) (27). This development extends far beyond early childhood, when children are mostly in the company and care of parents. Indeed, adolescence may be particularly important for supporting self-regulation because of the rapid growth, learning, adaptation, and neurobiological development that mark this period of life (28–30). Further, impulsive choices in adolescence (e.g., to start smoking, to drop out of school) can alter life trajectories in ways that are difficult to reverse (19).

Schools are a natural target for policy because of their potential to provide equal access to environments that support the development of self-regulation (31, 32). Not only is school where young people spend most of their waking hours outside the home, it is also where they experience a multitude of factors that have been shown to either scaffold or stymie the development of self-regulation, including adult role models (33, 34) and peers (35, 36). Recently, a growing chorus of policymakers has urged schools to extend their purview beyond traditional academic skills and into the domain of social-emotional skills such as self-regulation—a trend that is reflected in the expanded scope of federal and state standards and accountability systems (37–39).

In this investigation, we identify a pervasive measurement bias that, if not remedied, may thwart policymakers' efforts to evaluate, measure, and improve the effectiveness of schools that foster adolescent self-regulation. The possibility of this measurement bias has led to serious questions from policymakers about "whether we can make [self-regulation skills] visible, comparable, and therefore amenable to deliberate policy action in a similar way that traditional tests do with academic knowledge and skills" (40). As a result, education systems have been left with great interest in self-regulation and related constructs but insufficient scientific guidance.

The empirical starting point for our research is the mixed and often counterintuitive evidence regarding school effects on self-regulation. On the one hand, Jackson et al. (41) show encouraging evidence that schools can differ in how much they improve students scores on a self-report measure of hard work, and these school differences predicted students' later college enrollment and persistence. On the other hand, evaluations of charter schools show that they fail to raise self-reports of self-regulation, despite raising report card grades, standardized test scores, attendance rates, and college enrollment levels while reducing incarceration and unplanned pregnancies (42–45). Are high-performing schools whose cultures explicitly emphasize hard work and high expectations (46, 47) in fact having no impact on students' self-regulation—or is there a problem in how self-regulation is measured?

We suggest that reference bias, the systematic error that arises when respondents refer to different implicit standards when answering the same questions (48), is a legitimate threat to between-school comparisons, and could help explain the conflicting evidence of school effects on self-regulation. And even within a school, comparisons of students might be biased in the same way if different subgroups of students rely on different standards when answering the same questions. Reference bias can be especially pernicious in the present policy context because it can be difficult to detect and diagnose. Unlike social desirability bias and faking, reference bias can emerge even when respondents answer truthfully, and it can coexist with otherwise strong validity associations at the individual level. This is because reference bias could distort inferences any time there are comparisons of self-regulation across *groups* who differ in their frames of references—for example, schools with very different peer cultures with respect to effort, or even subcultures within a school.

Why might self-report questionnaires be subject to reference bias? Dominant models in survey methodology identify a multi-stage cog-

The authors declare no competing interests

Fig. 1. Peers influence the standards by which an individual judges their own behavior, resulting in a "reference bias" effect that distorts cross-context comparisons of self-reported self-regulation.

nitive response process: Students first read and interpret the question; then they identify relevant information in memory, form a summary judgment, and translate this judgment into one of the response options; finally, they edit their response if motivated to do so (49–51). As illustrated in **Figure 1,** a student may interpret a questionnaire item and its response options differently depending on their peers' typical behaviors (52). If they have high-achieving classmates who, for example, study for hours each evening and consistently arrive prepared for class, they might judge themselves against higher standards and rate themselves lower in self-regulation than an equally industrious student whose lower-achieving peers study and prepare less. While schools might be effective in increasing self-regulated behavior, they might at the same time increase the standards, leading to lower self-reported self-regulation.

To date, however, there has not been direct evidence that reference bias distorts comparisons between schools or any other groups. A handful of cross-cultural studies of self-regulation have yielded paradoxical findings (e.g., Asian countries like Japan and South Korea ranking lower in self-reported conscientiousness than other countries that are typically thought to be less conscientious (53)), but none of these studies directly measured the standard for behavior in the different contexts, relying instead on experts' ratings of cultural stereotypes or unvalidated proxies for self-regulation (e.g., the average walking speed in downtown locations of a convenience sample of a country's residents, as a proxy for the nation's conscientiousness).

In the educational literature more specifically, studies that compare the test scores and average self-regulation scores for different schools have not ruled out unobserved confounds, such as the possibility that school factors (e.g., average family income) that increase test scores (e.g., due to investment in educational opportunities) also decrease self-regulation (e.g., by shielding children from responsibilities that could cultivate self-regulation). Therefore, the research literature to date has not been able to distinguish biases in self-reports from potentially true group differences in self-regulation.

In this investigation, we overcome these limitations by using three complementary methods to examine reference bias more directly than has been possible previously. Our approach is motivated by the basic finding that people judge themselves compared to salient and similar others (52). Therefore we exploit (Studies 1 and 2) or work around (Study 3) variation in people's reference groups.

In Study 1 (total N = 206,589 students in k = 562 high schools), we show that the reference bias effect appears even within the same school in a year-over-year comparison. When students are surrounded by higher-achieving peers relative to other students at the same school in a different year, they rate themselves lower in self-regulation. Study 2 addresses an additional confound that could remain in Study 1's analysis, which is the possibility that year-over-year fluctuations in test scores are not random but are due to choices made by families about the academic trajectory of the school. In Study 2 (N = 21,818 students in k = 62 U.S. secondary schools), we rule this out with an analysis rooted in the purported psychological explanation for reference bias, which is that people's self-judgments should be more influenced by the peers whose behaviors they observe rather than peers whose behaviors they do not observe. We show that reference bias is evident in data from a single school year only when administrative data showed that the peers shared classes and therefore had an opportunity to observe each other's self-regulated behavior. Furthermore, Study 2 examined the theorized, but typically unmeasured, explanation for reference bias: differences in students' implicit standards for self-regulation.

Studies 1 and 2 ruled out school-level alternative explanations for reference bias but nevertheless allowed for the possibility that high-achieving peers reduce a student's real capacity for self-regulation.

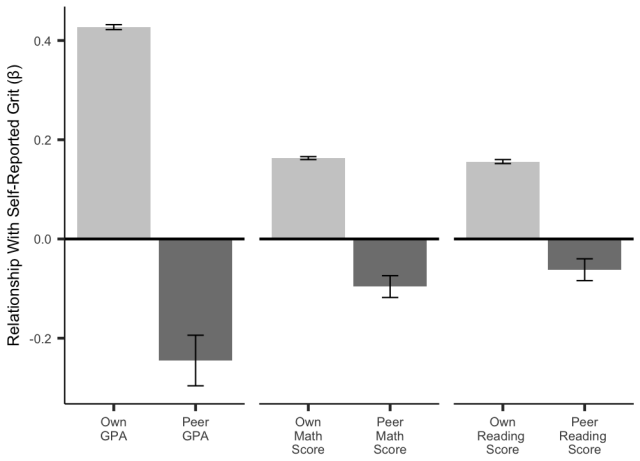

Fig. 2. In Study 1, self-reported grit correlated positively with a student's own academic performance but inversely with the performance of their schoolmates. OLS models included demographic controls and school fixed effects. Error bars represent ± 1 clustered standard errors.

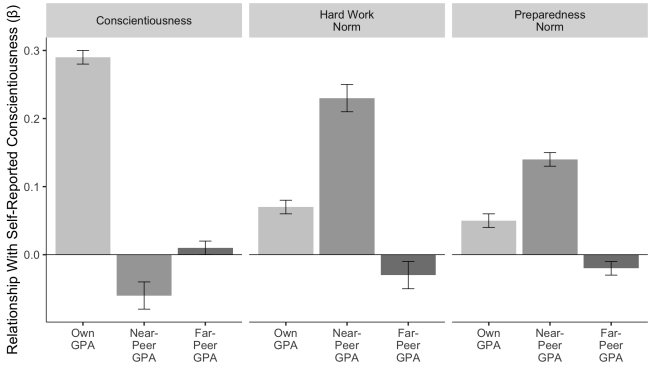

Fig. 3. In Study 2, self-reported conscientiousness correlated positively with a student's own GPA and negatively with the GPA of near-peers. In contrast, standards for what constitutes hard work and preparedness correlated positively with both own and near-peer GPA. As expected, there was no effect of far-peer GPA. OLS models included demographic controls and school fixed effects. Error bars represent ± 1 clustered standard errors.

Study 3 (N = 1,278 seniors in k = 15 U.S. high schools) addressed this possibility with a workaround: an objective behavioral task that involves no self-reports and therefore is not subject to biases due to differences in frames of reference. By matching self-regulation data collected in high school to records of college graduation, we show that there is no evidence of reference bias when a behavioral task is used. This evidence is bolstered by Study 3's use of a measure of school achievement that is independent of the high school peer group: graduation from college within 6 years after high school completion.

Results

Study 1: Evidence for reference bias in a country-wide natural experiment. In 2012 and 2013, the Secretariat of Public Education administered questionnaires measuring grit (the passion and perseverance for long term-term goals (54)) and collected data on academic performance from high school seniors in a nationally representative sample of 10% of high schools in Mexico. We analyzed data from the 1% of all schools that, by chance, were selected in both years. This enabled us to exploit exogenous variation in the academic performance of the 2013 high school cohort when compared to the performance of the 2012 cohort. Reference bias was quantified as the effect on self-reported grit uniquely attributable to peer academic performance (i.e., the cohort-wide averages of GPA, standardized math test scores, and standardized reading test scores, respectively, excluding said student from the average), after controlling for differences between schools, cohort year, and each student's own academic performance.

Students surrounded by higher-performing classmates rate themselves lower in grit. Consistent with prior research, among students in the same school, self-reported grit correlated positively with GPA (β = .43, p < .001), standardized math test scores (β = .16, p < .001), and standardized reading test scores (β = .16, p < .001). However, consistent with reference bias, self-reported grit correlated inversely with schoolmates' GPA (β = -.25, p < .001), peer standardized math test scores (β = -.09, p < .001), and peer standardized reading test scores (β = -.07, p = .004). See **Figure 2** and **Supporting Information** for details.

Evidence for reference bias was consistent across demographic subgroups. Capitalizing on the size and representativeness of our sample, we explored moderators of reference bias. Regression coefficients for peer academic performance were comparable whether subgroups were defined by gender, mother's educational level, school type (public or private), or school size. See **Supporting Information** for details.

Study 2: Replication and extension in a single large school district. In Study 2, we partnered with the nonprofit organization Character Lab to replicate and extend Study 1 with a sample of students in grades 8 through 12 in a large, diverse school district in the United States. This partnership enabled us to obtain official class schedules for each student, which we used to distinguish near- versus far-peers as students who did or didn't share daily academic classes, respectively. Whereas GPA was self-reported in Study 1, in Study 2 we obtained GPA from official school records. As part of a larger survey administered by Character Lab, students completed a self-report questionnaire of conscientiousness (the tendency to be organized, responsible, and hardworking (55)) as well as two questions we developed to directly assess self-regulation standards.

Reference bias replicates: Students whose classmates perform better academically rate themselves as lower in conscientiousness. As expected, this effect is driven by near-peers rather than far-peers. If implicit standards for self-regulation are determined by social comparison, reference bias should be driven by the individuals with whom individuals are in direct contact. As shown in **Figure 3,** consistent with Study 1, self-reported conscientiousness was correlated positively with a student's own GPA (β = .29, p < .001), negatively with the GPA of near-peers (β = - .06, p < .001), and not at all with the GPA of far-peers (β = .01, p = .395).

Students whose near-peers perform better academically hold higher self-regulation standards. As expected, standards for hard work were predicted by a student's own GPA (β = .07, p < .001) and the GPA of their near-peers (β = .23, p < .001), but not the GPA of their far-peers (β = -.03, p = .198). The same pattern emerged for preparedness norms, which were predicted by students own GPA (β = .05, p < .001) and the GPA of their near-peers (β = .14, p < .001), but not far-peers (β = -.02, p = .080). As in Study 1, the patterns of findings were generally similar across subgroups. See **Supporting Information** for details.

Study 3: In a longitudinal study of college graduation, evidence of reference bias in questionnaire but not task measures of self-regulation. In Study 3, we sought evidence of discriminant validity. Unlike questionnaires, which require participants to make subjective judgments of their behavior, task measures assay behavior directly. In a prospective, longitudinal study of N = 1,278 students attending k = 15 different college-preparatory charter schools in the United States, we tested the prediction that reference bias should be evident in questionnaire but not behavioral task measures of self-regulation. In their senior year of high school, students self-reported their grit and self-control (the ability to be in command of one's behavior and to inhibit one's impulses (55)). In addition, they completed the Academic Diligence Task, a behavioral task in which students voluntarily allocate attention to either good-for-me-later math problems or fun-for-me-now games and videos. The Academic Diligence Task has previously been validated as indexing self-control and grit (56, 57). Six years later, we used the National Student Clearinghouse database to identify students who successfully obtained their college diploma.

Evidence of reference bias in longitudinal predictions of college graduation from self-reported, but not objectively measured, self-regulation. As shown in **Figure 4,** among seniors in the same high school, higher scores on self-report questionnaires of self-control (b = 0.16, OR = 1.17, p = .022) and grit (b = 0.16, OR = 1.18, p = .020) each predicted greater odds of earning a college diploma 6 years later. However, college graduation rates were actually lower for schools with higher self-reported self-control and grit scores (b = -0.44, OR = 0.64, p = .001; b = -0.39, OR = 0.68, p = .005, for self-control and grit, respectively).

This paradoxical pattern was not evident when self-regulation was assessed objectively using the Academic Diligence Task (56). Among seniors in the same school, college graduation was predicted by higher scores on the Academic Diligence Task (b = 0.15, OR = 1.17, p = .031). Likewise, when comparing across schools, college graduation rates were higher for schools whose students performed better on the Academic Diligence Task (b = 0.46, OR = 1.58, p < .001). See **Supporting Information** for summaries of multilevel logistic regression models, robustness checks, and a replication of the own versus peer performance models in Studies 1 and 2.

Discussion

The three studies in this investigation provide direct evidence for reference bias in self-reported self-regulation. In Study 1, high school seniors rated themselves lower in grit when their schoolmates earned higher GPAs and standardized achievement test scores. In Study 2, we replicated this effect using self-report questionnaires of conscientiousness and showed that it was driven by near-peers rather than by far-peers. Further, we showed that the GPA of near-peers (but not far-peers) correlates positively with self-regulation standards (i.e., how many hours of homework constitute "a lot of homework" and how often it means to "sometimes" forget what they need for class). Finally, in Study 3, we found that using self-report questionnaires of grit and self-control to predict college graduation 6 years later produced paradoxical results: Within a high school, students with higher self-reported self-regulation were more likely to graduate from college 6 years later, but across schools, average levels of self-regulation negatively predicted graduation. In contrast, an objective task measure of self-regulation–which indexed performance directly and did not ask students to judge themselves–positively predicted college graduation both within and across schools.

How big are reference bias effects? Studies 1 and 2 provide estimates of r = .10 and .20. All else being equal, a student in our samples whose peers' academic achievement is one standard deviation above the mean is predicted to rate their own self-regulation as 10-20% of a standard deviation lower. Assuming that higher standards for self-regulation depress self-report ratings while at the same time, via social norms and modeling, encourage more self-regulated behavior, these are actually lower-bound estimates. Consistent with this possibility, when we use a behavioral task to assess self-regulation, we observe results consistent with positive peer effects (Study 3), which have also been previously reported in the literature (58–60). Taken together, our findings suggest that reference bias effects, even across social groups in the same country, can be at least small-to-medium in size by contemporary benchmarks (61) and comparable to the effect sizes for faking on self-regulation questionnaires in workplace settings (62).

Several limitations of the current investigation suggest promising directions for future research.

First, we must be cautious about drawing strong causal inferences from the non-experimental data in our three field studies. In Study 1, variation in peer quality could have influenced self-reported self-regulation for reasons other than reference bias. Against this, we found direct evidence for near-peer influence on self-regulation standards provided in Study 2. However, in Study 2, there is the possibility of reverse-causality. For example, rather than near-peers determining self-regulation standards, it is possible that self-regulation standards determined patterns of enrollment (e.g., students with higher standards self-selecting into the same difficult classes). In Study 3, we cannot rule out the possibility that some unmeasured confound gave rise to contradictory within-school versus between-school results on self-report (but not objective task) measures of self-regulation. In sum, it is important to confirm our observational findings by experimentally manipulating peer groups and/or standards of self-regulation.

Second, there are limits to the external validity of our conclusions. In particular, we examined reference bias in adolescence, a developmental period in which sensitivity to peers is at its apogee (63). The adolescents in our investigation lived in Mexico (Study 1) and the United States. (Studies 2 and 3). Further research on children and adults, in a wider sample of countries, and in contexts outside formal schooling, is needed to establish boundary conditions and moderators of reference bias. In general, effect sizes for reference bias are expected to be smaller when comparing social groups with more similar standards. Relatedly, while we found evidence for reference bias using three widely-used self-report questionnaires of self-regulation, we cannot assume that effects generalize to other questionnaires. For example, reference bias may have less impact on *inherently subjective* measures or those that are less observable in peers, such as self-esteem, academic self-concept, or life satisfaction (64, 65). These unexplored possibilities suggest that the effect sizes reported here might be different across samples and questionnaires.

Third, we did not collect nuanced data on social networks (e.g., friendships, acquaintances). Indeed, our operationalization of peer groups was quite crude–students in the same grade and attending the same school in Study 1 and 3, and students in the same grade and school who share at least one academic class (i.e., near-peers) in Study 2. Given the increasing prevalence of social-network studies and the continued popularity of self-report questionnaires in behavioral science, it should be possible to identify the influence of prominent social referents and close friends on reference bias.

Finally, while we collected information about student's standards for self-regulation (in Study 2) and an objective measure of self-regulation (in Study 3), we have yet to collect both types of measures

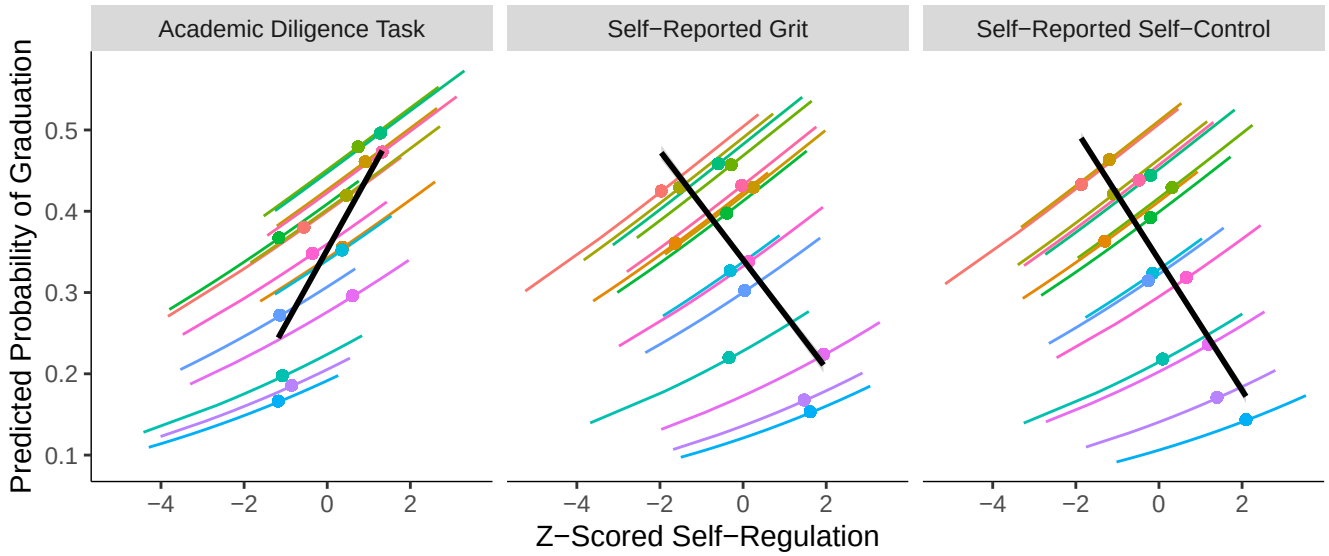

Fig. 4. In Study 3, comparing students within schools (colored lines), higher self-regulation predicted higher odds of college graduation, whether measured by self-report questionnaires for grit and self-control or by a behavioral task called the Academic Diligence Task. When comparing schools to each other, however, higher self-reported grit and self-control scores predicted *lower* graduation rates, whereas the behavioral task *positively* predicted college graduation, as shown in the solid black lines. Plots show predicted probabilities of graduation from multilevel logistic regression models.

in the same sample. Doing so in a future study would enable us to test a mediation model in which peers influence standards for self-regulation which, in turn, diminish self-reported self-regulation relative to performance on a behavioral task of self-regulation.

Unfortunately, the problem of reference bias is not easily corrected. The most commonly suggested solution is anchoring vignettes (66). This technique entails asking participants to rate detailed descriptions of hypothetical characters. These ratings are then used to adjust self-report questionnaire scores upward or downward depending on the stringency or leniency with which participants evaluated the hypothetical characters. Anchoring vignettes can increase the reliability and validity of self-reports (67) but do not always work as intended (68). They also increase the time, effort, and literacy required from survey respondents, which may limit their utility at scale (67, 69).

A related possibility is to use behaviorally anchored (70) or act-frequency rating scales (71), which ask respondents to rate themselves on more specific, contextualized behaviors than is typical in traditional questionnaires. For example, while students at over-subscribed charter schools do not rate themselves as more self-regulated, they and their parents do report more "minutes of homework completed" in an open-ended question in the same questionnaire (45). In our view, such questions might mitigate response bias but probably don't eliminate it altogether. Why not? Because all subjective judgments rely, at least to some degree, on implicit standards that can differ (e.g., What level of effort is sufficient to consider yourself to be "doing homework"?).

As shown in Study 3, self-regulation can be assessed with behavioral tasks, which appear immune to reference bias. However, task measures have their own limitations, including a dramatically lower signal-to-noise ratio when compared to questionnaires and, relatedly, surprisingly modest associations with other measures of self-regulation (51, 72–74).

Perhaps the best means of obviating reference bias is to take a multi-method, multi-informant approach to assessment, including trained observers who can rate behavior across multiple occasions (19). Observers who have seen hundreds, if not thousands, of cases typically have a wider reference frame than the individuals they are evaluating, which might explain why teacher ratings of behavior are more reliable and predictive of future outcomes than either parental reports or student self reports (75). The rarity of multi-method and multi-informant approaches suggests that, unfortunately, few researchers have the necessary resources or expertise to implement it, particularly at scale.

What are the implications of reference bias for researchers and policymakers?

Reference bias could suppress, or even reverse, the measured effects of interventions if the standards by which people judge their own behavior on pre- and post-questionnaires shift as a function of the intervention (76). In one study, participants were asked to rate their interviewing skills before training (*pre*). Afterward, participants rated themselves again (*post*) and, in addition, retrospectively estimated what their skills had been at baseline (*then*). Even though questionnaire items were identical for all assessments, *then* ratings were lower than *pre* ratings–suggesting that participants adopted higher standards as a result of the intervention. Moreover, third-party judges' ratings of performance matched *then-post* change better than *pre-post* differences (77).

The implications of reference bias extend beyond intervention research. Consider, for example, mean-level increases in conscientiousness across the lifespan (78–80). If adults in their 50s hold higher standards for what it means to be courteous, rule-abiding, and self-controlled than, say, teenagers, then age differences in conscientiousness may be even larger than we now think. In fact, to the extent that implicit standards and actual behavior are inversely correlated, reference bias should be expected to attenuate associations of self-regulation with groups of any kind.

While the importance of personal qualities like self-regulation is incontrovertible, the specter of reference bias argues against relying on self-report questionnaires when comparing students attending different schools, citizens who live in different countries, or indeed any the members of any social group whose standards could differ from one another. Are you a hard worker? Responding to such a question requires looking *within* to identify the patterns of our behavior.

In addition, the evidence for reference bias presented here suggests that knowingly or not, we also look *around* when we decide how to respond.

Methods

Study 1.

Sample and procedure. High school seniors in two representative random samples, each comprising 10% of schools in Mexico, completed standardized achievement tests of math and reading and, separately, self-report questionnaires late in the spring term of the 2011-12 and 2012-13 academic years, respectively. By chance, about 1% (k = 562) of high schools were included in both years. Our final sample includes 97.8% of the students in these high schools (N = 206,589) who completed a questionnaire measure of grit. Our sample was slightly biased towards girls (53.49% female). On average, students in our sample were 17.61 years old (SD = 0.79).

Self-reported grit. The Technical Committee for Background Questionnaires at the National Center of Evaluation for Higher Education in Mexico (Centro Nacional de Evaluación para la Educación Superior) translated all 8 items of the Short Grit Scale (81) as well as its 5-point Likert-type response scale (1 = *Not at all like me* to 5 = *Very much like me*) into Spanish. The observed reliability was $\alpha = .62$.

Grade point average (GPA). Students reported their overall GPA using a categorical scale ranging from *less than 5.9* to *10* in half-point increments (i.e., <5.9, 6.0-6.4, 6.5-6.9, etc.). We used the mid-point of the range in our analyses (i.e., 5.7, 6.2, 6.7, etc.). Although official GPAs were not available, meta-analytic estimates of the correlation between self-reported and objectively recorded GPA is $r = .82$ (82).

Standardized test scores. The Mexican Secretariat of Public Education provided standardized math and reading scores.

Analytic strategy. We used ordinary least squares (OLS) regression with clustered standard errors to predict self-reported grit from student's own and peer's academic performance:

$$G_{ist} = \alpha a_{ist} + \gamma b_{-ist} + \theta_s + \eta_t + \varepsilon_i$$

where G_{ist} is the self-reported grit for student i who was in 12th grade in school s at time t (2012 or 2013). Term a_{ist} is that student's own academic performance, operationalized as self-reported GPA, standardized math scores, or standardized verbal scores, respectively. Term b_{-ist} represents the average academic performance of students sharing a school with each student i, excluding student i. Term θ_s represents fixed effects for each student's school and captures ways in which schools might differ from each other—including such differences as teachers, curricula, school policies, and regional populations from which schools draw their members. Term η_t (fixed effect for year), captures how cohorts for each school systematically differ from each other. ε_i represents error.

Study 2.

Sample and procedure. This study included data from N = 21,818 (50% female, M_{age} = 15.60, SD_{age} = 1.54) students attending k = 62 middle and high schools in a large public school district in the United States who completed surveys in either October 2019 or February 2020. This district was part of Character Lab Research Network (CLRN), a consortium of school partners committed to advancing scientific insights that help children thrive. According to school records, the race/ethnicity of our sample was: Hispanic/Latinx (41%), White (28%), Black (23%), and other (8%). About half (49%) of students were eligible for free and reduced-price meals.

Self-reported conscientiousness. Students completed 12 items from the Big Five Inventory-2 (83) assessing conscientiousness (e.g., "I am someone who is persistent, works until the task is finished") using a 5-point Likert-type scale ranging from 1 = *Not like me at all* to 5 = *Totally like me*. The observed reliability was $\alpha = .83$.

Standards for hard work and preparedness. We included two questions to measure implicit standards for self-regulation. One question assessed norms for hard work: "If a student in your grade says they did 'a lot of homework' on a weeknight, how long would you guess they mean?" Eight response options ranged from 15 minutes (coded as 0.25 hours) to 3 or more hours (coded as 3 hours). The second question assessed norms for preparedness: "If a student in your grade says they 'sometimes' forget something they need for class, how often would you guess they mean?" Seven response options ranged from *once a month* to *three times or more per day* (coded as 66 times per month). We reverse-coded these values such that higher numbers indicated stricter standards for preparedness.

Grade point average (GPA). From school administrative records, we calculated GPAs on a 100-point scale by averaging final grades in students' academic courses (English language arts, math, science, social studies) for the quarter in which students took the survey during the 2019-2020 school year.

Near-peer and far-peer GPAs. For each student, we designated near-peers as those students who took at least one academic course with the target student during the quarter in which they took the survey. We designated far-peers as students in the same school who did *not* share any academic courses. For the average student in our sample, 38% of schoolmates were near-peers and 62% were far-peers.

Analytic strategy. To examine whether self-regulation standards and conscientiousness related to students' own and peers' performance, we fit OLS regression models with standard errors clustered by school to estimate the following equation:

$$S_{is} = \alpha a_{is} + \gamma_1 b_{-is} + \gamma_2 c_{-is} + \delta x_{is} + \theta_s + \varepsilon_i,$$

where S_{is} is a survey measure of conscientiousness or self-regulation standards for student i in school s, a_{is} is a student's own GPA, b_{-is} is the average GPA of students in the same school sharing at least one academic course with student i, c_{-is} is the average GPA of students in the same school but not sharing any academic courses with student i, x_{is} is a vector of student characteristics (age, gender, race/ethnicity, grade level, free or reduced-price meal status, English language learner status, special-education status, home language, and timing of the survey), θ_s represents school fixed effects, and ε_i is is a random error term.

Study 3.

Sample and procedure. A few weeks before graduation, N = 1,278 (55% female, M_{age} = 18.01, SD_{age} = 1.01) high school seniors responded to self-report questionnaires and task measures in school computer labs. Students attended k = 15 charter schools located in various urban centers in the United States. Between 76% and 98% of the students at each school participated in the study. Most students were socioeconomically disadvantaged (84% of students' mothers had less than a 4-year degree, 68% qualified for free or reduced-priced meals), and were mostly Latinx (46%) and African American (40%).

Self-reported grit. Students completed a 4-item version of the Grit Scale developed specifically for adolescents (84). Students responded on a 5-point Likert-type scale ranging from 1 = *Not at all true* to 5 = *Completely true*. The observed reliability was $\alpha = .78$.

Self-control. Students completed four items from the Domain-Specific Impulsivity Scale (56, 85) assessing academic self-control (e.g., "I forgot something needed for school"). Students responded on a 5-point Likert-type scale ranging from *Not at all true* to *Completely true.* The observed reliability was $\alpha = .72$.

Academic Diligence Task (ADT). A subset (n = 802) of students in our sample completed the Academic Diligence Task, a behavioral assessment of self-regulation that has been validated in separate research (56). This computer-based task begins with screens explaining that practicing simple mathematical skills like subtraction can aid in further enhancing overall math abilities. Then, they completed three 3-minute timed task blocks. In each, they chose between "Do math" and "Play game or watch movie." Clicking "Do math" displayed a math task involving single-digit subtraction with multiple-choice responses. On the other hand, clicking "Play game or watch movie" allowed students to play Tetris or watch entertaining videos. Participants could freely switch between them during each block. See **Supporting Information** for details. The key metric from the ADT was the mean number of problems correctly answered over the three blocks. Basic subtraction is very easy for most 12th

grade students, so attentive engagement with the task resulted almost exclusively in correct answers: The median rate of correct responses was 98.3%. Due to positive skew and some clustering of data at 0 (i.e., participants who did no math problems), we applied a square-root transformation to minimize bias from extremely high scores. This created an approximately normal distribution, which we used in subsequent calculations. Across the three blocks, the observed reliability was $\alpha = .78$.

General cognitive ability. During the online survey, students completed a brief (12-item) version of Raven's Progressive Matrices as an assessment of general cognitive ability (86). The ability variable was calculated as the sum of correctly answered questions out of 12, with any missing questions marked as incorrect. The observed reliability was $\alpha = .73$.

College graduation. Six years following data collection, we identified participants in the National Student Clearinghouse, a public database of college enrollment and graduation (87, 88). We coded for whether participants had obtained any college degree by that time.

Analytic strategy. Because students were clustered within schools, we used multilevel modeling to analyze how self-reported grit and self-control, and the Academic Diligence Task longitudinally predict graduation. Comparing the between-school coefficients for self-control and the ADT provides a test for reference bias. Specifically, we expect that while ADT will positively predict graduation both at the within- and between-school level, self-control will predict graduation within a school, but not between schools. We used a missing dummy variable coding approach to deal with missing data and included controls for general cognitive ability in our models.

ACKNOWLEDGMENTS. This research received support from the Bill & Melinda Gates Foundation, the Raikes Foundation, the William T. Grant Foundation, and a fellowship from the Center for Advanced Study in the Behavioral Sciences (CASBS) to the sixth author and grants from the John Templeton Foundation, the Walton Family Foundation, and National Science Foundation to the last author. This research was supported by the National Institute of Child Health and Human Development (Grant No. 10.13039/100000071 R01HD084772-01). The authors wish to thank Donald Kamentz, Laura Keane, and the schools and students who participated in the research.

1. Collaborative for Academic, Social, and Emotional Learning, What Is the CASEL Framework? (2021).
2. UNICEF, Measuring Life Skills: In the context of Life Skills and Citizenship Education in the Middle East and North Africa, Technical report (2020).
3. J. J. Heckman, Y. Rubinstein, The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *Am. Econ. Rev.* **91**, 145–149 (2001).
4. S. Lundberg, Noncognitive Skills as Human Capital in *Education, Skills, and Technological Change: Implications for Future US GDP Growth.* (University of Chicago Press, Chicago), pp. 219–250 (2019).
5. National Research Council, *Assessing 21st century skills: Summary of a workshop.* (The National Academies Press, Washington DC), (2011) Publisher: National Academies Press.
6. D. Park, E. Tsukayama, G. P. Goodwin, S. Patrick, A. L. Duckworth, A tripartite taxonomy of character: Evidence for intrapersonal, interpersonal, and intellectual competencies in children. *Contemp. Educ. Psychol.* **48**, 16–27 (2017).
7. A. Caspi, B. W. Roberts, R. L. Shiner, Personality Development: Stability and Change. *Annu. Rev. Psychol.* **56**, 453–484 (2005).
8. M. K. Rothbart, Temperament, Development, and Personality. *Curr. Dir. Psychol. Sci.* **16**, 207–212 (2007).
9. W. Mischel, Y. Shoda, M. L. Rodriguez, Delay of Gratification in Children. *Science* **244**, 933–938 (1989).
10. S. Freud, *Beyond the pleasure principle.*, Beyond the pleasure principle. eds. C. J. M. Hubback, E. Jones. (The International Psycho-Analytical Press, London, England), (1922) Pages: 90.
11. B. W. Roberts, H. J. Yoon, Personality Psychology. *Annu. Rev. Psychol.* **73**, 489–516 (2022).
12. J. T. Nigg, Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *J. Child Psychol. Psychiatry* **58**, 361–383 (2017).
13. G. S. Berns, D. Laibson, G. Loewenstein, Intertemporal choice – toward an integrative framework. *Trends Cogn. Sci.* **11**, 482–488 (2007).
14. J. J. Heckman, T. Kautz, Hard evidence on soft skills. *Labour Econ.* **19**, 451–464 (2012).
15. A. L. Duckworth, J. L. Taxer, L. Eskreis-Winkler, B. M. Galla, J. J. Gross, Self-control and academic achievement. *Annu. Rev. Psychol.* **70**, 373–399 (2019).
16. K. L. Bierman, R. L. Nix, M. T. Greenberg, C. Blair, C. E. Domitrovich, Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Dev. Psychopathol.* **20**, 821–843 (2008).
17. A. Vedel, The Big Five and tertiary academic performance: A systematic review and meta-analysis. *Pers. Individ. Differ.* **71**, 66–76 (2014).
18. M. Daly, M. Egan, J. Quigley, L. Delaney, R. F. Baumeister, Childhood self-control predicts smoking throughout life: Evidence from 21,000 cohort study participants. *Heal. Psychol.* **35**, 1254–1263 (2016).
19. T. E. Moffitt, et al., A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl. Acad. Sci.* **108**, 2693–2698 (2011).
20. T. Bogg, B. W. Roberts, Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychol. Bull.* **130**, 887–919 (2004).
21. W. Hofmann, M. Luhmann, R. R. Fisher, K. D. Vohs, R. F. Baumeister, Yes, But Are They Happy? Effects of Trait Self-Control on Affective Well-Being and Life Satisfaction: Trait Self-Control and Well-Being. *J. Pers.* **82**, 265–277 (2014).
22. T. Hirschi, Self-control and crime in *Handbook of self-regulation.* pp. 537–552 (year?).
23. M. R. Barrick, M. K. Mount, The big five personality dimensions and job performance: A meta-analysis. *Pers. Psychol.* **44**, 1–26 (1991).
24. A. L. Duckworth, D. Weir, E. Tsukayama, D. Kwok, Who Does Well in Life? Conscientious Adults Excel in Both Objective and Subjective Success. *Front. Psychol.* **3** (2012).
25. U. J. Wiersma, R. Kappe, Selecting for extroversion but rewarding for conscientiousness. *Eur. J. Work. Organ. Psychol.* **26**, 314–323 (2017).
26. J. J. A. Denissen, et al., Uncovering the power of personality to shape income. *Psychol. Sci.* **29**, 3–13 (2018).
27. S. Doebel, Rethinking Executive Function and Its Development. *Perspectives on Psychol. Sci.* **15**, 942–956 (2020).
28. B. J. Casey, Beyond Simple Models of Self-Control to Circuit-Based Accounts of Adolescent Behavior. *Annu. Rev. Psychol.* **66**, 295–319 (2015).
29. R. E. Dahl, N. B. Allen, L. Wilbrecht, A. B. Suleiman, Importance of investing in adolescence from a developmental science perspective. *Nature* **554**, 441–450 (2018).
30. L. Steinberg, Cognitive and affective development in adolescence. *Trends Cogn. Sci.* **9**, 69–74 (2005).
31. R. Bailey, E. A. Meland, G. Brion-Meisels, S. M. Jones, Getting Developmental Science Back Into Schools: Can What We Know About Self-Regulation Help Change How We Think About "No Excuses"? *Front. Psychol.* **10**, 1885 (2019).
32. S. F. Hamilton, Chapter 6: The Secondary School in the Ecology of Adolescent Development. *Rev. Res. Educ.* **11**, 227–258 (1984).
33. J. A. Leonard, Y. Lee, L. E. Schulz, Infants make more attempts to achieve a goal when they see adults persist. *Science* **357**, 1290–1294 (2017).
34. A. Bandura, W. Mischel, Modification of Self-Imposed delay of reward through exposure to live and suymbolic models. *J. Pers. Soc. Psychol.* **2**, 698–705 (1965).
35. K. M. King, K. A. McLaughlin, J. Silk, K. C. Monahan, Peer effects on self-regulation in adolescence depend on the nature and quality of the peer interaction. *Dev. Psychopathol.* **30**, 1389–1401 (2018).
36. S. Doebel, Y. Munakata, Group Influences on Engaging Self-Control: Children Delay Gratification and Value It More When Their In-Group Delays and Their Out-Group Doesn't. *Psychol. Sci.* **29**, 738–748 (2018).
37. J. P. Bertling, T. Marksteiner, P. C. Kyllonen, General Noncognitive Outcomes in *Assessing Contexts of Learning*, eds. S. Kuger, E. Klieme, N. Jude, D. Kaplan. (Springer International Publishing, Cham), pp. 255–281 (2016) Series Title: Methodology of Educational Measurement and Assessment.
38. U.S. Department of Education, Every Student Succeeds Act (ESSA) (2015).
39. Center for Disease Control and Prevention, Whole School, Whole Community, Whole Child (WSCC) (2021).
40. OECD, *Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills.* (OECD, Paris), (2021).
41. C. K. Jackson, S. C. Porter, J. Q. Easton, A. Blanchard, S. Kiguel, School effects on socioemotional development, school-based arrests, and educational attainment. *Am. Econ. Rev. Insights* **2**, 491–508 (2020).
42. M. R. West, et al., Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educ. Eval. Policy Analysis* **38**, 148–170 (2016).
43. W. Dobbie, R. G. Fryer, The medium-term impacts of high-achieving charter schools. *J. Polit. Econ.* **123**, 985–1037 (2015).
44. C. C. Tuttle, et al., Understanding the Effect of KIPP as it Scales: Volume I, Impacts on Achievement and Other Outcomes, (Mathematica Policy Research), Technical report (2015).
45. C. C. Tuttle, et al., KIPP Middle Schools: Impacts on Achievement and Other Outcomes, (Mathematica Policy Research), Technical report (2013).
46. J. D. Angrist, P. A. Pathak, C. R. Walters, Explaining Charter School Effectiveness. *Am. Econ. Journal: Appl. Econ.* **5**, 1–27 (2013).
47. W. Dobbie, R. G. Fryer, Getting Beneath the Veil of Effective Schools: Evidence From New York City. *Am. Econ. Journal: Appl. Econ.* **5**, 28–60 (2013).
48. S. J. Heine, D. R. Lehman, K. Peng, J. Greenholtz, What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *J. Pers. Soc. Psychol.* **82**, 903–918 (2002).
49. N. Schwarz, D. Oyserman, Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. *Am. J. Eval.* **22**, 127–160 (2001).
50. R. Tourangeau, L. J. Rips, K. Rasinski, *The psychology of survey response.* (Cambridge University Press, Cambridge), (2000).
51. A. L. Duckworth, D. S. Yeager, Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educ. Res.* **44**, 237–251 (2015).
52. N. Morina, Comparisons Inform Me Who I Am: A General Comparative-Processing Model of Self-Perception. *Perspectives on Psychol. Sci.* **16**, 1281– 1299 (2021).
53. D. P. Schmitt, J. Allik, R. R. McCrae, V. Benet-Martínez, The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *J. Cross-Cultural Psychol.* **38**, 173–212 (2007).
54. A. L. Duckworth, C. Peterson, M. D. Matthews, D. R. Kelly, Grit: Perseverance and passion for long-term goals. *J. Pers. Soc. Psychol.* **92**, 1087–1101 (2007).
55. American Psychological Association, *APA Dictionary of Psychology.* (American Psychological Association, Washington DC), 1 edition, (2007).
56. B. M. Galla, et al., The Academic Diligence Task (ADT): Assessing individual differences in effort on tedious but important schoolwork. *Contemp. Educ. Psychol.* **39**, 314–325 (2014).
57. G. Zamarro, M. Nichols, A. Duckworth, S. D'Mello, Further Validation of Survey Effort Measures of Relevant Character Skills: Results from a Sample of High School Students. *EDRE Work. Pap.* **2018-07** (2018).

58. R. B. Cialdini, Descriptive social norms as underappreciated sources of social control. *Psychometrika* **72**, 263–268 (2007).
59. A. Bandura, *Social learning theory.* (General Learning Press, New York), (1971).
60. B. Sacerdote, Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? in *Handbook of the Economics of Education*. (Elsevier) Vol. 3, pp. 249–277 (2011).
61. D. C. Funder, D. J. Ozer, Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Adv. Methods Pract. Psychol. Sci.* **2**, 156–168 (2019).
62. A. Martínez, J. F. Salgado, A Meta-Analysis of the Faking Resistance of Forced-Choice Personality Inventories. *Front. Psychol.* **12**, 732241 (2021).
63. L. Steinberg, S. M. Amanda, Adolescent development. *Annu. Rev. Psychol.* **52**, 83–110 (2000).
64. S. Oishi, D. P. Roth, The role of self-reports in culture and personality research: It is too early to give up on self-reports. *J. Res. Pers.* **43**, 107–109 (2009).
65. H. W. Marsh, The Big-Fish-Little-Pond effect on academic self-concept. *J. Educ. Psychol.* **79**, 280–295 (1987).
66. G. King, C. J. L. Murray, J. A. Salomon, A. Tandon, Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am. Polit. Sci. Rev.* **98**, 191–207 (2004).
67. R. Primi, C. Zanon, D. Santos, F. De Fruyt, O. P. John, Anchoring Vignettes: Can They Make Adolescent Self-Reports of Social-Emotional Skills More Reliable, Discriminant, and Criterion-Valid? *Eur. J. Psychol. Assess.* **32**, 39–51 (2016).
68. H. Grol-Prokopczyk, E. Verdes-Tennant, M. McEniry, M. Ispány, Promises and Pitfalls of Anchoring Vignettes in Health Survey Research. *Demography* **52**, 1703–1728 (2015).
69. J. P. Bertling, F. Borgonovi, D. E. Almonte, Psychosocial Skills in Large-Scale Assessments: Trends, Challenges, and Policy Implications in *Psychosocial Skills and School Systems in the 21st Century: Theory, Research, and Practice*, The Springer Series on Human Exceptionality, eds. A. A. Lipnevich, F. Preckel, R. D. Roberts. (Springer International Publishing, Switzerland), (2016).
70. D. P. Schwab, H. G. Heneman, T. A. DeCotiis, Behaviorally anchored rating scales: A review of the literature. *Pers. Psychol.* **28**, 549–562 (1975).
71. D. M. Buss, K. H. Craik, The Act Frequency Approach to Personality. *Psychol. Rev.* **90**, 105–126 (1983).
72. A. Z. Enkavi, et al., Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proc. Natl. Acad. Sci.* **116**, 5472–5477 (2019).
73. A. L. Duckworth, M. L. Kern, A meta-analysis of the convergent validity of self-control measures. *J. Res. Pers.* **45**, 259–268 (2011).
74. L. Sharma, K. Kohl, T. A. Morgan, L. A. Clark, "Impulsivity": Relations between self-report and behavior. *J. Pers. Soc. Psychol.* **104**, 559–575 (2013).
75. S. Feng, Y. Han, J. J. Heckman, T. Kautz, Comparing the reliability and predictive power of child, teacher, and guardian reports of noncognitive skills. *Proc. Natl. Acad. Sci.* **119**, e2113992119 (2022).
76. G. S. Howard, Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Eval. Rev.* **4**, 93–106 (1980).
77. G. S. Howard, P. R. Dailey, Response-Shift Bias: A source of contamination of self-report measures. *J. Appl. Psychol.* **64**, 144–150 (1979).
78. B. W. Roberts, K. E. Walton, W. Viechtbauer, Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychol. Bull.* **132**, 1–25 (2006).
79. R. I. Damian, M. Spengler, Sixteen going on sixty-six: A longitudinal study of personality stability and change across 50 years. *J. Pers. Soc. Psychol.* **117**, 674–695 (2018).
80. B. W. Roberts, W. F. DelVecchio, The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychol. Bull.* **126**, 3–25 (2000).
81. A. L. Duckworth, P. D. Quinn, Development and validation of the Short Grit scale (Grit–S). *J. Pers. Assess.* **91**, 166–174 (2009).
82. N. R. Kuncel, M. Credé, L. L. Thomas, The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Rev. Educ. Res.* **75**, 63–82 (2005).
83. C. J. Soto, O. P. John, Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *J. Res. Pers.* **68**, 69–81 (2017).
84. B. M. Galla, et al., Why high school grades are better predictors of on-time college graduation than are admissions test scores: The roles of self-regulation and cognitive ability. *Am. Educ. Res. J.* **56**, 2077–2115 (2019) Publisher: American Educational Research Association.
85. E. Tsukayama, A. L. Duckworth, B. Kim, Domain-specific impulsivity in school-age children. *Dev. Sci.* **16**, 879–893 (2013).
86. J. Raven, J. Raven, Raven progressive matrices in *Handbook of nonverbal assessment*, ed. R. S. McCallum. (Kluwer Academic/Plenum Publishers, New York), pp. 223–237 (2003).
87. S. M. Dynarski, S. W. Hemelt, J. M. Hyman, The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes. *Educ. Eval. Policy Analysis* **37**, 53S–79S (2015).
88. C. Schoenecker, R. Reeves, The National Student Clearinghouse: The largest current student tracking database. *New Dir. for Community Coll.* **143**, 47–57 (2008).

Supplementary Information for

Large Studies Reveal How Reference Bias Limits Policy Applications of Self-Report Measures

Benjamin Lira*, Joseph M. O'Brien*, Pablo A. Peña, Brian M. Galla, Sidney D'Mello, David S. Yeager, Amy Defnet, Tim Kautz, Kate Munkacsy, Angela L. Duckworth

Benjamin Lira.
E-mail: blira@upenn.edu

This PDF file includes:

Other supplementary materials for this manuscript include the following:

List of Figures

List of Tables

Supporting Information Text

1. Study 1

Table S1. Demographic variables included as controls in OLS models in Study 1.

Question
Gender
How old are you?
Marital status
Does your mother speak an indigenous language or dialect?
Does your father speak an indigenous language or dialect?
Do you live with your mother?
Do you live with your father?
When you have an academic doubt, do you have someone to help you?
What is the level of education reached by your mother (even if she passed away)?
What is the level of education reached by your father (even if he passed away)?
How many books are there in your house? (Don't include magazines, newspapers, or textbooks.)
In the last 2 years, how many times have you gone on vacation within the Mexican Republic?
How many states of the Mexican Republic have you visited as a tourist?
At home, do you have a conventional telephone (landline)?
At home, do you have a cell phone?
At home, do you have a washing machine?
At home, do you have a refrigerator?
At home, do you have a gas stove?
At home, do you have a microwave oven?
At home, do you have cable or satellite TV?
At home, do you have Internet?
How many DVD players are there in your house?
How many computers are there in your house?
How many televisions are there in your house?
How many cars are there in your house?

Table S2. Descriptive statistics, bivariate correlations, and partial correlations for Study 1

	1	2	3	4	5	6	7	8	9	10	11	12
Own												
1. Grit		0.28***	0.25***	0.25***	0.13***	0.15***	0.12***	−0.02***	−0.01***	−0.03***	0.01**	0.01***
2. Overall GPA	0.26***		0.61***	0.57***	0.36***	0.36***	−0.01***	0.03***	0.02***	0.02***	0.00*	0.00
3. Math GPA	0.24***	0.65***		0.43***	0.42***	0.29***	0.00*	0.02***	0.04***	0.01***	0.01*	0.00
4. Verbal GPA	0.25***	0.61***	0.46***		0.20***	0.30***	−0.02***	0.02***	0.01***	0.04***	0.01***	0.01***
5. Math test score	0.14***	0.29***	0.36***	0.19***		0.53***	0.01**	0.01**	0.01***	0.01***	0.13***	0.07***
6. Reading test score	0.16***	0.31***	0.27***	0.29***	0.63***		0.01***	0.00	0.00	0.01***	0.07***	0.11***
Peers'												
7. Grit	0.21***	0.05***	0.05***	0.08***	0.12***	0.14***		0.01***	0.04***	−0.09***	0.08***	0.12***
8. Overall GPA	0.03***	0.45***	0.31***	0.30***	0.04***	0.07***	0.13***		0.58***	0.59***	0.18***	0.15***
9. Math GPA	0.03***	0.37***	0.36***	0.26***	0.08***	0.09***	0.14***	0.83***		0.37***	0.20***	0.12***
10. Verbal GPA	0.05***	0.35***	0.24***	0.39***	0.10***	0.14***	0.22***	0.77***	0.65***		0.16***	0.22***
11. Math test score	0.05***	0.04***	0.05***	0.07***	0.55***	0.46***	0.22***	0.08***	0.15***	0.18***		0.59***
12. Reading test score	0.06***	0.06***	0.07***	0.11***	0.49***	0.51***	0.28***	0.14***	0.19***	0.27***	0.89***	
M	3.72	8.38	8.05	8.70	0.10	0.10	3.72	8.38	8.05	8.70	0.10	0.10
SD	0.56	0.73	1.11	0.85	1.00	1.01	0.13	0.33	0.42	0.34	0.56	0.52
N	206,589	210,423	201,445	190,658	211,255	211,255	206,589	210,423	201,445	190,658	211,255	211,255

Notes. Bivariate correlations under diagonal. Partial correlations controlling for school dummy variables shown above the diagonal. Missing data was handled with pairwise deletion.

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

Table S3. OLS regression models predicting self-reported grit from self and peer academic performance in Study 1

	Overall GPA	Math GPA	Language GPA	Math Test Score	Reading Test Score
Own performance	0.427***	0.233***	0.308***	0.163***	0.156***
	(0.005)	(0.004)	(0.004)	(0.003)	(0.004)
Peer performance	−0.245***	−0.133***	−0.196***	−0.096***	−0.062**
	(0.051)	(0.028)	(0.039)	(0.022)	(0.022)
R^2	0.124	0.110	0.109	0.071	0.071
N	205,901	197,139	186,551	206,589	206,589

Notes. All regressions include high school fixed effects, year dummies, and controls for student characteristics. Robust standard errors clustered by high school. The number of observations differs across regressions because of missing values.

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

Table S4. Subgroup analyses by range of individual GPA in Study 1

Individual GPA Range	Peer-Average GPA	Peer-Average Language GPA	Peer-Average Math GPA
<6		0.599	0.211
		(1.250)	(0.176)
6.0 - 6.4	0.744	−0.511	−0.101
	(0.855)	(0.374)	(0.067)
6.5 - 6.9	0.193	−0.016	−0.093
	(0.303)	(0.288)	(0.093)
7.0 - 7.4	−0.132	−0.250*	−0.162**
	(0.132)	(0.124)	(0.061)
7.5 - 7.9	−0.172	−0.158	−0.136**
	(0.088)	(0.109)	(0.065)
8.0 - 8.4	−0.224**	−0.140**	−0.141**
	(0.074)	(0.058)	(0.051)
8.5 - 8.9	−0.312***	−0.229**	−0.125**
	(0.088)	(0.072)	(0.062)
9.0 - 9.4	−0.459***	−0.280***	−0.324***
	(0.102)	(0.075)	(0.063)
9.5 - 9.9	−0.126	−0.241	−0.112
	(0.153)	(0.130)	(0.114)
10	−0.937	−0.125	0.095
	(0.948)	(0.118)	(0.113)

Notes. Models additionally include parameters for individuals own GPA, controls for student characteristics, and fixed effects for school and class year, not shown. Robust standard errors clustered by high school.

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

Table S5. Subgroup analyses by individual performance quintiles in Study 1

Individual Test Score Quintile	Math Test Score	Reading Test Score
Bottom 20%	−0.062	−0.037
	(0.043)	(0.039)
Quintile 2	−0.031	−0.040
	(0.041)	(0.042)
Quintile 3	−0.106**	−0.064
	(0.040)	(0.039)
Quintile 4	−0.130**	−0.033
	(0.041)	(0.040)
Top 20%	−0.102*	−0.100
	(0.050)	(0.054)

Notes. Models additionally include parameters for individuals own test scores, controls for student characteristics, and fixed effects for school and class year, not shown. Robust standard errors clustered by high school.

† p < .10, * p < .05, ** p < .01, *** p < .001.

Table S6. Subgroup analyses by gender, school type, and mother's education in Study 1

	Overall GPA		Math GPA		Language GPA		Math Test Score		Reading Test Score	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Student gender: Boys (1) vs. Girls (2)										
Own performance	0.390***	0.466***	0.222***	0.246***	0.277***	0.342***	0.141***	0.185***	0.135***	0.181***
	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)
Peer performance	−0.207**	−0.275***	−0.146***	−0.124***	−0.226***	−0.175***	−0.063**	−0.135***	−0.055	−0.076**
	(0.066)	(0.061)	(0.039)	(0.034)	(0.051)	(0.051)	(0.029)	(0.029)	(0.029)	(0.029)
R^2	0.112	0.139	0.102	0.121	0.099	0.122	0.064	0.080	0.064	0.080
n	95,447	110,400	92,472	104,616	87,218	99,282	95,780	110,755	95,780	110,755
School type: Public (1) vs. Private (2)										
Own performance	0.421***	0.454***	0.231***	0.241***	0.302***	0.338***	0.161***	0.171***	0.156***	0.157***
	(0.005)	(0.011)	(0.004)	(0.007)	(0.004)	(0.009)	(0.004)	(0.008)	(0.004)	(0.006)
Peer performance	−0.255***	−0.242**	−0.132***	−0.096	−0.182***	−0.206**	−0.101***	−0.078	−0.056*	−0.130**
	(0.059)	(0.089)	(0.031)	(0.059)	(0.043)	(0.076)	(0.024)	(0.046)	(0.024)	(0.044)
R^2	0.123	0.135	0.111	0.113	0.108	0.115	0.072	0.073	0.072	0.073
n	170,188	35,713	162,280	34,859	153,070	33,481	170,769	35,820	170,769	35,820
Mothers education: less than high school (1) vs. more than high school (2)										
Own performance	0.424***	0.428***	0.236***	0.229***	0.304***	0.311***	0.175***	0.149***	0.174***	0.138***
	(0.006)	(0.006)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Peer performance	−0.207***	−0.251***	−0.072*	−0.186***	−0.162***	−0.214***	−0.109***	−0.074***	−0.062*	−0.056*
	(0.062)	(0.071)	(0.035)	(0.037)	(0.049)	(0.052)	(0.025)	(0.031)	(0.028)	(0.029)
R^2	0.130	0.124	0.117	0.109	0.116	0.108	0.079	0.069	0.080	0.069
n	104,058	100,531	98,866	97,066	93,280	92,127	104,452	100,803	104,452	100,803

Notes. All regressions include high school fixed effects, year dummies, and controls for student characteristics not used in the subgroup analyses. Robust standard errors clustered by high school.

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

Table S7. Subgroup analyses by school size tertiles in Study 1

	Smallest third	Middle third	Largest third
Overall GPA			
Own performance	0.436***	0.438***	0.421***
	(0.012)	(0.008)	(0.006)
Peer performance	−0.152	−0.118	−0.357***
	(0.085)	(0.087)	(0.081)
R^2	0.150	0.130	0.115
n	19,957	57,676	128,268
Math GPA			
Own performance	0.258***	0.251***	0.222***
	(0.008)	(0.005)	(0.005)
Peer performance	−0.060	−0.126**	−0.167***
	(0.055)	(0.047)	(0.044)
R^2	0.138	0.116	0.101
n	19,023	54,254	123,862
Language GPA			
Own performance	0.328***	0.325***	0.298***
	(0.011)	(0.007)	(0.005)
Peer performance	−0.121	−0.138	−0.270***
	(0.071)	(0.074)	(0.058)
R^2	0.131	0.115	0.099
n	18,250	51,242	117,059
Math Test Scores			
Own performance	0.202***	0.176***	0.151***
	(0.010)	(0.006)	(0.004)
Peer performance	−0.158***	−0.090*	−0.086*
	(0.040)	(0.035)	(0.037)
R^2	0.094	0.073	0.063
n	20,057	57,950	128,582
Reading Test Scores			
Own performance	0.191***	0.175***	0.143***
	(0.010)	(0.006)	(0.005)
Peer performance	−0.156**	−0.033	−0.054
	(0.046)	(0.035)	(0.032)
R^2	0.0795	0.075	0.063
n	20,057	57,950	128,582

Notes. Models additionally include parameters for individuals own GPA, controls for student characteristics, and fixed effects for school and class year, not shown. Robust standard errors clustered by high school.

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

2. Study 2

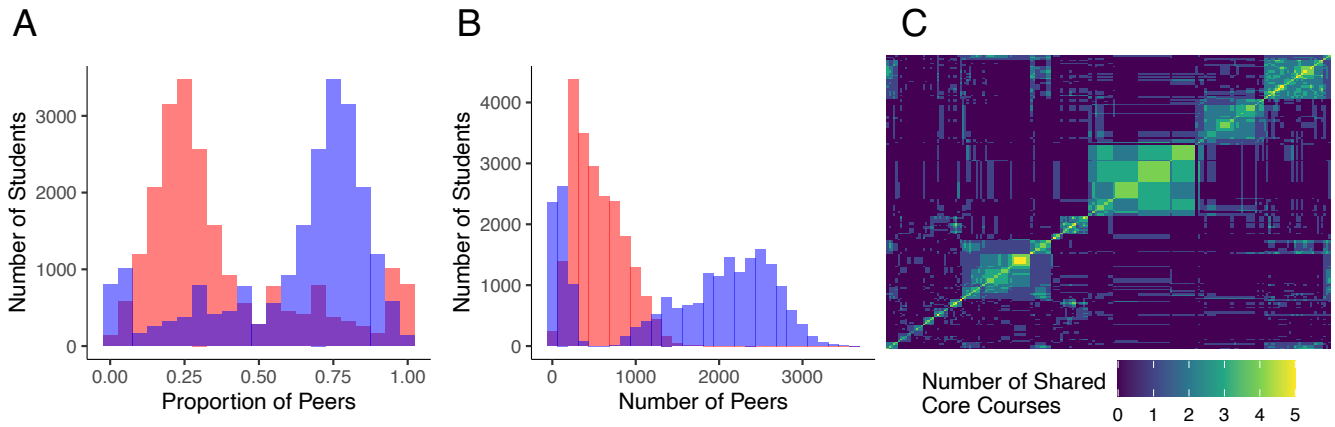

Fig. S1. Details about the construction of the near-peer and far-peer GPA variables. A. Histograms of the proportion of peers within each school that constituted near (blue) and far (red) peer variables. B. Histograms of the absoolute number of peers within each school that constituted near (blue) and far (red) peer variables. C. Example similarity matrix depicting the number of shared core courses between any two students. Each colored square is a pair of students. The diagonal represents the number of core courses each student is taking. Lighter colors represent more shared core courses.

Table S8. Means, standard deviations, and correlation coefficients among the study variables in Study 2

	1	2	3	4	5	6	7
Own							
1. Preparedness norm	-	0.15***	0.06***	0.11***	-	0.15***	−0.10***
2. Hard work norm	0.18***	-	0.09***	0.19***	-	0.25***	−0.14***
3. Conscientiousness	0.06***	0.10***	-	0.25***	-	0.05***	−0.02***
4. Core GPA	0.13***	0.19***	0.24***	-	-	0.42***	−0.28***
Peer							
5. Core GPA (in same school)	0.09***	0.10***	−0.02**	0.22***	-	-	-
6. Core GPA (peers in shared core courses)	0.17***	0.25***	0.03***	0.46***	0.62***	-	−0.54***
7. Core GPA (peers not in shared core courses)	−0.02**	−0.07***	−0.05***	−0.08***	0.55***	0.02***	-
Student demographics							
Female	0.03***	0.12***	0.09***	0.15***	−0.02**	0.04***	−0.04***
Age	0.03***	0.12***	0.09***	−0.04***	−0.10***	0.00	−0.16***
English language learner	−0.04***	−0.05***	0.03***	−0.12***	−0.04***	−0.26***	0.15***
Special education student	0.02*	0.03***	−0.03***	0.03***	0.04***	0.05***	0.04***
Eligible for free or reduced-price meals	−0.12***	−0.16***	−0.01	−0.24***	−0.19***	−0.30***	−0.02*
Race/ethnicity							
Hispanic	−0.05***	−0.06***	−0.02**	−0.11***	0.09***	−0.10***	0.17***
Black, non-Hispanic	−0.09***	−0.10***	0.03***	−0.13***	−0.35***	−0.25***	−0.21***
White, non-Hispanic	0.12***	0.12***	−0.01	0.16***	0.20***	0.27***	0.03***
Other	0.04***	0.07***	0.01	0.13***	0.05***	0.13***	−0.03***
Grade level							
8	−0.06***	−0.17***	−0.07***	0.04***	0.07***	0.07***	0.18***
9	0.03***	0.02*	−0.02***	−0.01	0.05***	−0.10***	0.02*
10	−0.02**	0.01	−0.01	−0.06***	−0.10***	−0.11***	−0.09***
11	0.03***	0.10***	0.05***	0.00	0.00	0.05***	−0.06***
12	0.03***	0.07***	0.06***	0.04***	−0.04***	0.09***	−0.08***
Home language							
English	0.06***	0.05***	0.00	0.06***	0.04***	0.15***	−0.07***
Spanish	−0.06***	−0.06***	0.00	−0.09***	0.03***	−0.14***	0.13***
Other	0.00	0.00	0.01	0.03***	−0.10***	−0.04***	−0.08***
Took survey in winter term (versus fall term)	−0.03***	−0.08***	−0.07***	−0.02**	−0.16***	−0.08***	−0.08***
M	50.13	1.90	3.39	80.97	80.10	80.01	80.60
SD	18.82	0.93	0.62	10.36	2.76	4.18	4.20
N	21,818	21,818	21,818	21,818	21,818	21,818	21,818

Notes. Bivariate correlations appear below the diagonal. Partial Correlations controlling for school dummy variables appear above the diagonal. We do not show partial correlations between peer GPA and other variables because the peer terms are collinear with the school dummy indicators.
* p < .05, ** p < .01, *** p < .001.

Table S9. Differences across schools and intraclass correlation coefficients in Study 2

		Not controlling for student demographics		Controlling for student demographics		
	Kurtosis	R^2	F	R^2	F	ICC
Conscientiousness	1.083	0.015	5.562***	0.030	8.905***	0.013
Hard work Norm	−0.525	0.079	30.733***	0.118	38.361***	0.077
Preparedness Norm	−0.598	0.030	10.904***	0.043	12.724***	0.026
Big Five Conscientiousness items						
Reliable	−0.688	0.010	3.475***	0.038	11.421***	0.008
Irresponsible	−0.587	0.012	4.155***	0.029	8.661***	0.010
Dependable	−0.308	0.009	3.294***	0.021	6.272***	0.007
Careless	−0.177	0.006	2.127***	0.014	4.201***	0.003
Difficulty starting	0.129	0.016	5.824***	0.023	6.633***	0.012
Lazy	0.729	0.011	4.065***	0.021	6.190***	0.010
Tidy	2.007	0.012	4.380***	0.029	8.624***	0.009
Systematic	2.033	0.008	3.002***	0.021	6.225***	0.006
Messy	2.302	0.018	6.473***	0.023	6.793***	0.014
Persistent	2.676	0.009	3.147***	0.017	4.829***	0.007
Disorganized	2.725	0.008	2.910***	0.016	4.693***	0.005
Efficient	4.526	0.007	2.612***	0.014	4.167***	0.005

Notes. Bootstrapped confidence intervals with B = 1000 replications show that there is significantly more between-school variation for both standards (95% CIs [0.071 – 0.096] and [0.025 – 0.041], for hard work and preparedness norms respectively) for self-regulation compared to conscientiousness (95% CI [0.013 – 0.023]).

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

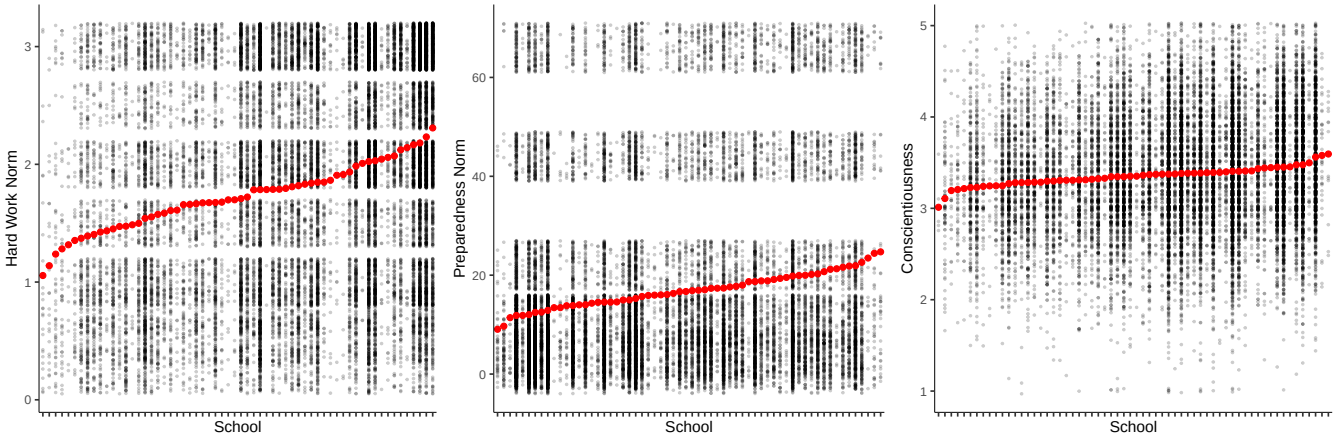

Fig. S2. Within- and between-school variation in conscientiousness and self-regulation standards. Each black dot is an individual student, arranged in vertical stacks, each of which represents a school. The red dots are the school-wide means for each variable.

Table S10. OLS regression models predicting self-reported conscientiousness and self-regulation standards, respectively, in Study 2

	Conscientiousness					Hard work norm					Preparedness norm				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Own GPA	0.258***	0.288***	0.258***	0.288***	0.292***	0.183***	0.097***	0.183***	0.097***	0.071***	0.113***	0.058***	0.113***	0.058***	0.045***
	(0.007)	(0.007)	(0.012)	(0.012)	(0.013)	(0.007)	(0.007)	(0.011)	(0.008)	(0.007)	(0.007)	(0.008)	(0.011)	(0.009)	(0.009)
School-peer GPA	0.082**		0.082			−0.015		−0.015			−0.015		−0.015		
	(0.030)		(0.041)			(0.030)		(0.046)			(0.031)		(0.033)		
Near-peer GPA		−0.065***		−0.065***	−0.059***		0.248***		0.248***	0.230***		0.150***		0.150***	0.143***
		(0.011)		(0.015)	(0.016)		(0.010)		(0.025)	(0.024)		(0.011)		(0.013)	(0.013)
Far-Peer GPA		0.038**		0.038*	0.012		−0.011		−0.011	−0.025		−0.023		−0.023*	−0.021
		(0.012)		(0.016)	(0.014)		(0.011)		(0.020)	(0.020)		(0.012)		(0.011)	(0.012)
Race/ethnicity															
White, non-Hispanic (reference)															
Black, non-Hispanic					0.034***					−0.030**					−0.053***
					(0.009)					(0.009)					(0.007)
Hispanic					0.018					−0.016					−0.035***
					(0.011)					(0.010)					(0.009)
Other Race					0.001					0.007					−0.011
					(0.008)					(0.007)					(0.007)
Grade level															
Grade 8 (reference)															
Grade 9					−0.030					−0.029					0.059
					(0.077)					(0.024)					(0.073)
Grade 10					−0.041					0.024					0.041
					(0.073)					(0.023)					(0.067)
Grade 11					−0.034					0.080**					0.058
					(0.084)					(0.028)					(0.077)
Grade 12					−0.037					0.079**					0.045
					(0.070)					(0.026)					(0.063)
Student demographics															
Female					0.045***					0.096***					0.019**
					(0.007)					(0.007)					(0.007)
Eligible for free or reduced priced meals					0.024**					−0.043***					−0.032***
					(0.007)					(0.010)					(0.006)
English language learner					0.052***					0.036***					0.020*
					(0.008)					(0.007)					(0.009)
Special education student					−0.026***					0.008					−0.006
					(0.007)					(0.007)					(0.007)
Took survey in the winter term					−0.044***					−0.017					0.008
					(0.009)					(0.014)					(0.009)
Age					0.105***					−0.115***					−0.016
					(0.015)					(0.017)					(0.021)
Home language															
English (reference)															
Spanish					0.008					0.007					−0.015
					(0.012)					(0.010)					(0.010)
Other					−0.027*					−0.002					−0.0001
					(0.010)					(0.006)					(0.008)
N	21,818	21,818	21,818	21,818	21,818	21,818	21,818	21,818	21,818	21,818	21,818	21,818	21,818	21,818	21,818
R^2	0.079	0.082	0.079	0.082	0.095	0.111	0.143	0.111	0.143	0.159	0.042	0.055	0.042	0.055	0.059

Notes. For each variable we report 5 models. (1) School-wide peers and no student characteristics, (2) Near- and far-peers and no student characteristics, (3) School-wide peers, no student characteristics, and school fixed-effects, (4) Near- and far-peers, no student characteristics, and school fixed-effects, (5) Near- and far-peers, student characteristics, and school fixed-effects. Robust standard errors clustered by school.

Notes. * p < .05, ** p < .01, *** p < .001.

Table S11. Subgroup analyses by gender in Study 2

	Male			Female		
	C	HWN	PN	C	HWN	PN
Own GPA	0.288***	0.078***	0.057***	0.293***	0.066***	0.033**
	(0.013)	(0.011)	(0.013)	(0.015)	(0.010)	(0.010)
Near-peer GPA	−0.089***	0.187***	0.103***	−0.029	0.280***	0.187***
	(0.019)	(0.027)	(0.019)	(0.018)	(0.026)	(0.015)
Far-Peer GPA	0.011	−0.043*	−0.030	0.009	−0.009	−0.010
	(0.019)	(0.018)	(0.018)	(0.018)	(0.027)	(0.018)
Race/ethnicity						
White, non-Hispanic (reference)						
Black, non-Hispanic	0.041***	−0.038**	−0.052***	0.028	−0.021	−0.053***
	(0.010)	(0.013)	(0.014)	(0.016)	(0.013)	(0.012)
Hispanic	0.032	−0.015	−0.046**	0.003	−0.017	−0.026*
	(0.019)	(0.012)	(0.015)	(0.017)	(0.014)	(0.011)
Other Race	0.016	0.005	−0.021*	−0.015	0.007	−0.002
	(0.009)	(0.011)	(0.010)	(0.011)	(0.008)	(0.011)
Grade level						
Grade 8 (reference)						
Grade 9	−0.087	−0.052	0.126	0.030	−0.018	0.009
	(0.118)	(0.080)	(0.131)	(0.073)	(0.039)	(0.060)
Grade 10	−0.089	−0.009	0.108	0.007	0.043	−0.009
	(0.111)	(0.076)	(0.121)	(0.074)	(0.034)	(0.059)
Grade 11	−0.087	0.064	0.117	0.020	0.082	0.021
	(0.126)	(0.088)	(0.137)	(0.087)	(0.041)	(0.071)
Grade 12	−0.067	0.078	0.071	−0.003	0.070	0.037
	(0.105)	(0.072)	(0.113)	(0.074)	(0.041)	(0.062)
Student demographics						
Eligible for free or reduced priced meals	0.035***	−0.049**	−0.048***	0.014	−0.036***	−0.016
	(0.010)	(0.015)	(0.012)	(0.011)	(0.009)	(0.010)
English language learner	0.070***	0.040***	0.037***	0.037***	0.035**	0.004
	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)	(0.012)
Special education student	−0.034**	0.007	−0.006	−0.015	0.010	−0.005
	(0.010)	(0.008)	(0.009)	(0.010)	(0.012)	(0.009)
Took survey in the winter term	−0.036*	−0.029	0.005	−0.055***	−0.003	0.014
	(0.014)	(0.015)	(0.013)	(0.012)	(0.017)	(0.012)
Age	0.090**	−0.151***	0.029	0.120***	−0.081***	−0.063
	(0.028)	(0.025)	(0.021)	(0.026)	(0.022)	(0.034)
Home language						
English (reference)						
Spanish	0.004	−0.009	−0.010	0.014	0.024	−0.017
	(0.015)	(0.013)	(0.015)	(0.015)	(0.016)	(0.012)
Other	−0.025	0.005	−0.010	−0.031*	−0.009	0.011
	(0.013)	(0.008)	(0.010)	(0.013)	(0.009)	(0.008)
n	10,868	10,868	10,868	10,950	10,950	10,950
R^2	0.088	0.133	0.053	0.099	0.169	0.074

Notes Estimates calculated by running separate OLS regressions for each subgroup. Robust standard errors clustered by school. C = Conscientiousness, HWN = Hard work norm, PN = Preparedness norm.

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

Table S12. Subgroup analyses by grade in Study 2

	Grade 8			Grade 9			Grade 10			Grade 11			Grade 12		
	C	HWN	PN	C	HWN	PN	C	HWN	PN	C	HWN	PN	C	HWN	PN
Own GPA	0.334***	0.091***	0.051**	0.328***	0.076***	0.065*	0.300***	0.089***	0.042*	0.260***	0.044**	0.036	0.231***	0.043	0.029*
	(0.016)	(0.013)	(0.015)	(0.014)	(0.014)	(0.025)	(0.016)	(0.022)	(0.018)	(0.021)	(0.014)	(0.018)	(0.021)	(0.027)	(0.013)
Near-peer GPA	−0.080*	0.074	0.079**	−0.060*	0.203***	0.138***	−0.084*	0.102*	0.066*	−0.020	0.299***	0.187***	−0.055*	0.241***	0.162***
	(0.033)	(0.055)	(0.027)	(0.029)	(0.039)	(0.029)	(0.035)	(0.039)	(0.031)	(0.028)	(0.034)	(0.022)	(0.020)	(0.046)	(0.022)
Far-Peer GPA	0.027	−0.127**	−0.061*	−0.038	−0.148**	−0.076	−0.007	−0.325***	−0.193**	0.093	−0.097	0.026	−0.147	−0.015	−0.041
	(0.030)	(0.042)	(0.026)	(0.044)	(0.046)	(0.046)	(0.056)	(0.048)	(0.053)	(0.067)	(0.077)	(0.034)	(0.093)	(0.112)	(0.086)
Race/ethnicity															
White, non-Hispanic (reference)															
Black, non-Hispanic	0.041*	−0.034	−0.061***	0.036	−0.002	−0.044*	0.024	−0.066**	−0.056**	0.045*	−0.018	−0.049**	0.014	−0.034	−0.030
	(0.020)	(0.023)	(0.015)	(0.021)	(0.018)	(0.017)	(0.017)	(0.018)	(0.020)	(0.020)	(0.020)	(0.017)	(0.025)	(0.035)	(0.029)
Hispanic	0.011	−0.002	−0.035	0.001	0.006	−0.052*	0.012	−0.032	−0.033	0.037	−0.005	−0.015	0.033	−0.061***	−0.037
	(0.026)	(0.023)	(0.022)	(0.021)	(0.021)	(0.024)	(0.019)	(0.018)	(0.023)	(0.021)	(0.019)	(0.021)	(0.017)	(0.015)	(0.031)
Other Race	−0.0004	−0.009	−0.019	0.004	0.012	−0.004	−0.007	0.003	−0.029	−0.004	0.016	0.006	0.008	0.009	−0.011
	(0.018)	(0.015)	(0.013)	(0.008)	(0.017)	(0.018)	(0.018)	(0.017)	(0.019)	(0.021)	(0.014)	(0.010)	(0.015)	(0.022)	(0.017)
Student demographics															
Female	0.001	0.071***	0.0003	0.054***	0.076***	0.032	0.056*	0.128***	0.010	0.071***	0.112***	0.016	0.062***	0.103***	0.057**
	(0.011)	(0.012)	(0.011)	(0.014)	(0.015)	(0.016)	(0.020)	(0.015)	(0.019)	(0.012)	(0.010)	(0.015)	(0.013)	(0.026)	(0.017)
Eligible for free or reduced priced meals	−0.003	−0.047**	−0.027*	0.040**	−0.043*	−0.032*	0.030	−0.032	−0.035**	0.034*	−0.028	−0.033	0.039*	−0.046***	−0.013
	(0.011)	(0.014)	(0.013)	(0.011)	(0.019)	(0.011)	(0.016)	(0.022)	(0.011)	(0.012)	(0.015)	(0.017)	(0.018)	(0.012)	(0.022)
English language learner	0.046**	0.050***	0.036	0.078***	0.032	0.050*	0.042*	0.023	−0.010	0.062***	0.034*	0.015	0.028	0.040	−0.015
	(0.013)	(0.014)	(0.020)	(0.015)	(0.028)	(0.019)	(0.017)	(0.012)	(0.017)	(0.016)	(0.015)	(0.015)	(0.024)	(0.021)	(0.014)
Special education student	0.033*	0.019	−0.002	−0.051**	−0.004	−0.031*	−0.031**	0.002	0.004	−0.055***	0.029	0.013	−0.015	−0.047***	−0.036**
	(0.015)	(0.015)	(0.017)	(0.015)	(0.016)	(0.014)	(0.009)	(0.016)	(0.016)	(0.009)	(0.015)	(0.012)	(0.016)	(0.011)	(0.011)
Took survey in the winter term	−0.034	−0.019	0.025	−0.057**	−0.057***	−0.029	−0.035	−0.099***	−0.026	−0.030	−0.028	0.020	−0.069*	0.027	0.034
	(0.021)	(0.020)	(0.018)	(0.019)	(0.015)	(0.024)	(0.018)	(0.022)	(0.017)	(0.031)	(0.025)	(0.014)	(0.027)	(0.024)	(0.029)
Age	0.037***	−0.030*	−0.008	0.019	−0.042	0.030	0.073***	−0.040*	−0.001	0.031*	−0.046***	−0.024	0.048**	−0.059***	−0.011
	(0.010)	(0.014)	(0.012)	(0.019)	(0.023)	(0.021)	(0.017)	(0.015)	(0.024)	(0.012)	(0.009)	(0.015)	(0.013)	(0.014)	(0.016)
Home language															
English (reference)															
Spanish	0.049*	0.002	−0.002	0.006	0.016	−0.020	−0.026	−0.010	−0.042	−0.027	0.010	0.003	0.027	0.033	−0.029
	(0.019)	(0.024)	(0.018)	(0.023)	(0.021)	(0.024)	(0.020)	(0.028)	(0.030)	(0.017)	(0.019)	(0.019)	(0.025)	(0.019)	(0.031)
Other	0.008	−0.005	0.002	−0.031**	−0.021	0.002	−0.039	0.007	0.033	−0.047*	0.004	−0.017	−0.051*	0.006	−0.015
	(0.013)	(0.016)	(0.015)	(0.010)	(0.016)	(0.011)	(0.020)	(0.014)	(0.020)	(0.021)	(0.012)	(0.016)	(0.020)	(0.017)	(0.013)
n	5,903	5,903	5,903	4,177	4,177	4,177	3,606	3,606	3,606	5,091	5,091	5,091	3, 041	3, 041	3, 041
R^2	0.101	0.095	0.043	0.117	0.158	0.075	0.099	0.163	0.072	0.080	0.166	0.067	0.082	0.136	0.071

Notes. Estimates calculated by running separate OLS regressions for each subgroup. Robust standard errors clustered by school. C = Conscientiousness, HWN = Hard work norm, PN = Preparedness norm.

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

Table S13. Subgroup analyses by eligibility for free or reduced priced meals status in Study 2

	Eligible for free or reduced priced meals			Not eligible for free or reduced priced meals		
	C	HWN	PN	C	HWN	PN
Own GPA	0.270***	0.058***	0.024*	0.299***	0.082***	0.067***
	(0.013)	(0.009)	(0.011)	(0.015)	(0.010)	(0.012)
Near-peer GPA	−0.068***	0.214***	0.123***	−0.036*	0.235***	0.151***
	(0.019)	(0.025)	(0.018)	(0.016)	(0.028)	(0.017)
Far-Peer GPA	0.007	−0.028	−0.037*	0.028	−0.022	0.001
	(0.016)	(0.021)	(0.017)	(0.019)	(0.026)	(0.021)
Race/ethnicity						
White, non-Hispanic (reference)						
Black, non-Hispanic	0.051**	−0.042*	−0.045***	0.020	−0.015	−0.046***
	(0.018)	(0.016)	(0.012)	(0.010)	(0.012)	(0.013)
Hispanic	0.030	−0.029	−0.019	0.010	−0.006	−0.038**
	(0.018)	(0.016)	(0.014)	(0.013)	(0.013)	(0.011)
Other Race	0.008	0.015	−0.007	−0.006	0.001	−0.012
	(0.012)	(0.010)	(0.009)	(0.008)	(0.011)	(0.011)
Grade level						
Grade 8 (reference)						
Grade 9	0.133	−0.029	0.054	−0.141**	−0.032	0.075
	(0.118)	(0.055)	(0.055)	(0.044)	(0.050)	(0.107)
Grade 10	0.119	0.025	0.029	−0.141**	0.022	0.064
	(0.115)	(0.056)	(0.054)	(0.041)	(0.048)	(0.095)
Grade 11	0.152	0.082	0.049	−0.160**	0.073	0.081
	(0.126)	(0.063)	(0.062)	(0.051)	(0.056)	(0.115)
Grade 12	0.124	0.075	0.043	−0.144**	0.080	0.059
	(0.106)	(0.054)	(0.053)	(0.047)	(0.051)	(0.094)
Student demographics						
Female	0.024**	0.093***	0.025*	0.064***	0.100***	0.012
	(0.008)	(0.010)	(0.011)	(0.010)	(0.009)	(0.010)
English language learner	0.064***	0.026**	0.030**	0.035***	0.049**	0.005
	(0.011)	(0.009)	(0.010)	(0.010)	(0.015)	(0.014)
Special education student	−0.014	−0.016	−0.010	−0.039***	0.030**	−0.005
	(0.009)	(0.010)	(0.012)	(0.010)	(0.009)	(0.008)
Took survey in the winter term	−0.060***	−0.033*	0.012	−0.026*	0.001	0.008
	(0.012)	(0.015)	(0.012)	(0.012)	(0.015)	(0.012)
Age	0.102***	−0.120***	−0.008	0.102***	−0.108***	−0.025
	(0.021)	(0.023)	(0.031)	(0.026)	(0.025)	(0.025)
Home language						
English (reference)						
Spanish	−0.001	0.020	−0.015	0.018	−0.004	−0.017
	(0.014)	(0.014)	(0.012)	(0.013)	(0.012)	(0.013)
Other	−0.055***	−0.014	−0.004	0.004	0.010	0.009
	(0.013)	(0.010)	(0.011)	(0.010)	(0.009)	(0.010)
n	10,704	10,704	10,704	11,114	11,114	11,114
R^2	0.100	0.129	0.037	0.101	0.156	0.065

Notes. Estimates calculated by running separate OLS regressions for each subgroup. Robust standard errors clustered by school. C = Conscientiousness, HWN = Hard work norm, PN = Preparedness norm.

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

Table S14. Subgroup analyses by race/ethnicity in Study 2

	White			Black			Hispanic			Other Race		
	C	HWN	PN	C	HWN	PN	C	HWN	PN	C	HWN	PN
Own GPA	0.333***	0.076***	0.080***	0.209***	0.059***	−0.003	0.299***	0.074***	0.049***	0.246***	0.033	0.057
	(0.014)	(0.011)	(0.014)	(0.019)	(0.017)	(0.018)	(0.011)	(0.012)	(0.011)	(0.034)	(0.034)	(0.038)
Near-peer GPA	−0.001	0.265***	0.191***	−0.071**	0.214***	0.145***	−0.064**	0.203***	0.101***	−0.086	0.174***	0.141***
	(0.021)	(0.028)	(0.023)	(0.022)	(0.029)	(0.017)	(0.020)	(0.028)	(0.021)	(0.051)	(0.045)	(0.038)
Far-Peer GPA	0.071*	0.029	0.060*	−0.030	−0.014	−0.028	0.012	−0.058*	−0.055**	−0.037	−0.052	−0.025
	(0.030)	(0.026)	(0.028)	(0.023)	(0.025)	(0.018)	(0.019)	(0.026)	(0.017)	(0.048)	(0.048)	(0.051)
Grade level												
Grade 8 (reference)												
Grade 9	−0.019	0.313***	0.190	−0.036	0.022	−0.032	−0.102	−0.176***	−0.016	0.263	0.044	0.298
	(0.035)	(0.059)	(0.243)	(0.122)	(0.049)	(0.098)	(0.056)	(0.037)	(0.045)	(0.223)	(0.248)	(0.216)
Grade 10	−0.038	0.324***	0.159	−0.046	0.079	−0.053	−0.101	−0.123***	−0.031	0.227	0.123	0.229
	(0.024)	(0.050)	(0.195)	(0.147)	(0.058)	(0.119)	(0.052)	(0.034)	(0.044)	(0.208)	(0.230)	(0.199)
Grade 11	−0.066*	0.425***	0.178	−0.004	0.177**	−0.055	−0.096	−0.098*	0.001	0.259	0.186	0.285
	(0.029)	(0.071)	(0.245)	(0.155)	(0.064)	(0.132)	(0.061)	(0.042)	(0.050)	(0.225)	(0.266)	(0.229)
Grade 12	−0.083**	0.341***	0.145	−0.019	0.185**	−0.022	−0.072	−0.078*	−0.022	0.227	0.188	0.215
	(0.031)	(0.061)	(0.190)	(0.137)	(0.059)	(0.120)	(0.051)	(0.037)	(0.044)	(0.204)	(0.232)	(0.199)
Student demographics												
Female	0.073***	0.108***	0.014	0.048***	0.084***	0.009	0.025*	0.095***	0.017	0.032	0.112***	0.074**
	(0.012)	(0.011)	(0.014)	(0.013)	(0.013)	(0.011)	(0.010)	(0.009)	(0.009)	(0.021)	(0.024)	(0.024)
Eligible for free or reduced priced meals	0.011	−0.037**	−0.040**	0.021*	−0.034	−0.029*	0.031***	−0.046***	−0.016	−0.004	−0.023	−0.051
	(0.013)	(0.011)	(0.012)	(0.010)	(0.022)	(0.014)	(0.009)	(0.012)	(0.009)	(0.017)	(0.021)	(0.028)
English language learner	0.035**	−0.003	0.008	0.025	0.012	−0.011	0.072***	0.055***	0.029*	0.030	−0.019	−0.017
	(0.011)	(0.021)	(0.017)	(0.018)	(0.013)	(0.016)	(0.013)	(0.012)	(0.013)	(0.025)	(0.025)	(0.027)
Special education student	−0.056***	0.015	−0.001	−0.038*	0.002	0.018	0.0001	0.003	−0.025*	−0.011	0.021	−0.026
	(0.013)	(0.014)	(0.013)	(0.016)	(0.020)	(0.015)	(0.010)	(0.012)	(0.010)	(0.027)	(0.024)	(0.020)
Took survey in the winter term	0.010	0.039*	0.031	−0.057**	−0.034	−0.001	−0.066***	−0.041*	0.001	−0.043	−0.025	0.040
	(0.018)	(0.019)	(0.021)	(0.016)	(0.017)	(0.015)	(0.013)	(0.018)	(0.016)	(0.044)	(0.032)	(0.039)
Age	0.142***	−0.077	−0.032	0.069*	−0.187***	−0.012	0.107***	−0.078*	−0.016	0.073	−0.178**	0.010
	(0.027)	(0.044)	(0.030)	(0.026)	(0.033)	(0.051)	(0.022)	(0.031)	(0.030)	(0.047)	(0.059)	(0.060)
Home language												
English (reference)												
Spanish	0.018	0.013	0.002	−0.009	0.015	−0.003	0.008	0.003	−0.019	0.002	0.019	0.010
	(0.010)	(0.012)	(0.014)	(0.010)	(0.015)	(0.014)	(0.013)	(0.013)	(0.011)	(0.007)	(0.026)	(0.018)
Other	0.026	0.016	0.007	−0.075***	−0.005	−0.002	−0.005	0.003	0.018	−0.030	−0.0003	−0.010
	(0.014)	(0.012)	(0.012)	(0.016)	(0.012)	(0.018)	(0.008)	(0.010)	(0.010)	(0.027)	(0.024)	(0.019)
n	6,166	6,166	6,166	5,020	5,020	5,020	8,913	8,913	8,913	1,719	1,719	1,719
R^2	0.129	0.175	0.069	0.082	0.134	0.043	0.111	0.133	0.040	0.089	0.186	0.121

Notes. Estimates calculated by running separate OLS regressions for each subgroup. Robust standard errors clustered by school. C = Conscientiousness, HWN = Hard work norm, PN = Preparedness norm.

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

Table S15. Subgroup analyses by English language learner status in Study 2

	English Language Learner			Non English Language Learner		
	C	HWN	PN	C	HWN	PN
Own GPA	0.256***	0.127***	0.040	0.296***	0.064***	0.045***
	(0.028)	(0.023)	(0.021)	(0.014)	(0.008)	(0.010)
Near-peer GPA	−0.047	−0.034	−0.047	−0.058***	0.240***	0.150***
	(0.035)	(0.033)	(0.035)	(0.017)	(0.023)	(0.013)
Far-Peer GPA	0.046	−0.009	−0.120**	0.013	−0.035	−0.024
	(0.065)	(0.079)	(0.038)	(0.016)	(0.018)	(0.013)
Race/ethnicity						
White, non-Hispanic (reference)						
Black, non-Hispanic	−0.042	−0.041	−0.111*	0.037***	−0.028**	−0.047***
	(0.044)	(0.054)	(0.043)	(0.010)	(0.009)	(0.007)
Hispanic	−0.093**	0.018	0.020	0.022*	−0.019	−0.036***
	(0.034)	(0.051)	(0.047)	(0.011)	(0.010)	(0.009)
Other Race	−0.032	0.0001	−0.013	0.0002	0.007	−0.010
	(0.022)	(0.028)	(0.027)	(0.009)	(0.008)	(0.008)
Grade level						
Grade 8 (reference)						
Grade 9	0.002	−0.096	−0.013	−0.009	−0.037	0.061
	(0.045)	(0.049)	(0.035)	(0.094)	(0.022)	(0.078)
Grade 10	−0.067**	−0.067*	−0.061	−0.017	0.021	0.046
	(0.024)	(0.033)	(0.032)	(0.089)	(0.021)	(0.072)
Grade 11				−0.011	0.077**	0.057
	(0.000)	(0.000)	(0.000)	(0.103)	(0.026)	(0.082)
Grade 12	−0.031	0.040	−0.022	−0.015	0.076**	0.045
	(0.027)	(0.029)	(0.025)	(0.087)	(0.025)	(0.068)
Student demographics						
Female	−0.033	0.061**	−0.047*	0.052***	0.099***	0.026***
	(0.020)	(0.021)	(0.020)	(0.007)	(0.007)	(0.007)
Eligible for free or reduced priced meals	0.034	−0.072**	0.037	0.021*	−0.036***	−0.036***
	(0.017)	(0.023)	(0.020)	(0.008)	(0.010)	(0.007)
Special education student	0.015	0.004	−0.049	−0.028***	0.009	−0.003
	(0.015)	(0.023)	(0.027)	(0.007)	(0.008)	(0.007)
Took survey in the winter term	−0.119***	−0.101***	0.001	−0.036**	−0.009	0.006
	(0.022)	(0.024)	(0.024)	(0.011)	(0.013)	(0.010)
Age	0.150**	−0.063	−0.060	0.099***	−0.124***	−0.004
	(0.048)	(0.045)	(0.038)	(0.018)	(0.018)	(0.022)
Home language						
English (reference)						
Spanish	−0.023	−0.048	−0.129	0.003	0.007	−0.010
	(0.079)	(0.084)	(0.079)	(0.011)	(0.009)	(0.009)
Other	−0.128	−0.079	−0.045	−0.021	0.002	0.002
	(0.091)	(0.076)	(0.074)	(0.011)	(0.006)	(0.008)
n	2,091	2,091	2,091	19,727	19,727	19,727
R^2	0.124	0.086	0.059	0.097	0.171	0.064

Notes. Estimates calculated by running separate OLS regressions for each subgroup. Robust standard errors clustered by school. C = Conscientiousness, HWN = Hard work norm, PN = Preparedness norm.
*p < .05; **p < .01; *** < .001

Table S16. Subgroup analyses by GPA tertiles in Study 2

	Low GPA			Medium GPA			High GPA		
	C	HWN	PN	C	HWN	PN	C	HWN	PN
Own GPA	0.353***	0.076	0.147***	0.553***	0.132**	0.115***	0.159***	0.061***	0.036
	(0.060)	(0.054)	(0.040)	(0.046)	(0.047)	(0.029)	(0.021)	(0.015)	(0.022)
Near-peer GPA	−0.072***	0.273***	0.145***	−0.069**	0.174***	0.120***	−0.033	0.243***	0.149***
	(0.020)	(0.027)	(0.021)	(0.021)	(0.047)	(0.022)	(0.023)	(0.027)	(0.022)
Far-Peer GPA	0.023	0.012	−0.029	0.002	−0.030	0.009	0.040	−0.071*	−0.031
	(0.022)	(0.022)	(0.022)	(0.028)	(0.031)	(0.021)	(0.028)	(0.030)	(0.019)
Race/ethnicity									
White, non-Hispanic (reference)									
Black, non-Hispanic	0.032	−0.005	−0.062***	0.003	−0.029	−0.061***	0.083***	−0.047**	−0.034*
	(0.017)	(0.015)	(0.016)	(0.019)	(0.016)	(0.016)	(0.013)	(0.015)	(0.016)
Hispanic	0.033	0.003	−0.058***	0.0001	−0.007	−0.003	0.048**	−0.033	−0.028
	(0.021)	(0.015)	(0.015)	(0.019)	(0.016)	(0.017)	(0.017)	(0.021)	(0.014)
Other Race	0.014	0.008	−0.022*	−0.030**	0.002	−0.002	0.027	0.017	−0.018
	(0.017)	(0.015)	(0.010)	(0.009)	(0.010)	(0.011)	(0.016)	(0.020)	(0.018)
Grade level									
Grade 8 (reference)									
Grade 9	−0.026	−0.179*	−0.196***	−0.177**	0.024	0.141	0.249*	−0.070	0.316**
	(0.053)	(0.078)	(0.045)	(0.053)	(0.058)	(0.091)	(0.101)	(0.092)	(0.118)
Grade 10	−0.052	−0.123	−0.184***	−0.162**	0.094	0.105	0.227*	−0.028	0.287*
	(0.048)	(0.073)	(0.042)	(0.050)	(0.054)	(0.084)	(0.096)	(0.088)	(0.112)
Grade 11	−0.057	−0.103	−0.184***	−0.148*	0.148*	0.107	0.271*	0.039	0.347*
	(0.058)	(0.085)	(0.050)	(0.056)	(0.065)	(0.096)	(0.109)	(0.100)	(0.131)
Grade 12	−0.069	−0.082	−0.135**	−0.120*	0.135*	0.073	0.224*	0.049	0.280*
	(0.050)	(0.072)	(0.042)	(0.047)	(0.056)	(0.083)	(0.092)	(0.084)	(0.109)
Student demographics									
Female	0.045***	0.085***	0.022*	0.059***	0.112***	0.022	0.023	0.086***	0.011
	(0.012)	(0.010)	(0.011)	(0.011)	(0.012)	(0.011)	(0.012)	(0.011)	(0.012)
Eligible for free or reduced priced meals	0.039***	−0.031	−0.025	0.019	−0.060***	−0.036**	0.021	−0.038**	−0.027*
	(0.011)	(0.016)	(0.015)	(0.015)	(0.011)	(0.011)	(0.012)	(0.013)	(0.013)
English language learner	0.061***	0.058***	0.033*	0.026	−0.005	−0.014	0.057***	0.036*	0.019
	(0.014)	(0.012)	(0.013)	(0.017)	(0.017)	(0.019)	(0.011)	(0.014)	(0.013)
Special education student	−0.041**	0.002	−0.024*	−0.026**	0.029**	0.011	−0.036***	−0.011	−0.014
	(0.014)	(0.013)	(0.011)	(0.009)	(0.010)	(0.010)	(0.010)	(0.012)	(0.013)
Took survey in the winter term	−0.043*	−0.009	−0.004	−0.043*	0.001	−0.006	−0.051***	−0.041*	0.031*
	(0.018)	(0.013)	(0.016)	(0.020)	(0.020)	(0.013)	(0.014)	(0.018)	(0.015)
Age	0.155***	−0.093**	−0.062**	0.005	−0.104**	0.017	0.120***	−0.132***	−0.002
	(0.032)	(0.030)	(0.023)	(0.026)	(0.034)	(0.039)	(0.034)	(0.026)	(0.037)
Home language									
English (reference)									
Spanish	0.008	0.010	0.004	−0.001	−0.003	−0.050**	0.012	0.014	−0.007
	(0.018)	(0.013)	(0.015)	(0.020)	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)
Other	−0.029*	−0.012	−0.008	−0.025*	0.004	−0.010	−0.033*	0.0001	0.016
	(0.013)	(0.011)	(0.014)	(0.010)	(0.012)	(0.009)	(0.014)	(0.011)	(0.013)
n	7,243	7,243	7,243	7,061	7,061	7,061	7,514	7,514	7,514
R^2	0.061	0.138	0.062	0.053	0.155	0.069	0.071	0.128	0.038

Notes. Estimates calculated by running separate OLS regressions for each subgroup. Robust standard errors clustered by school. C = Conscientiousness, HWN = Hard work norm, PN = Preparedness norm.

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

3. Study 3

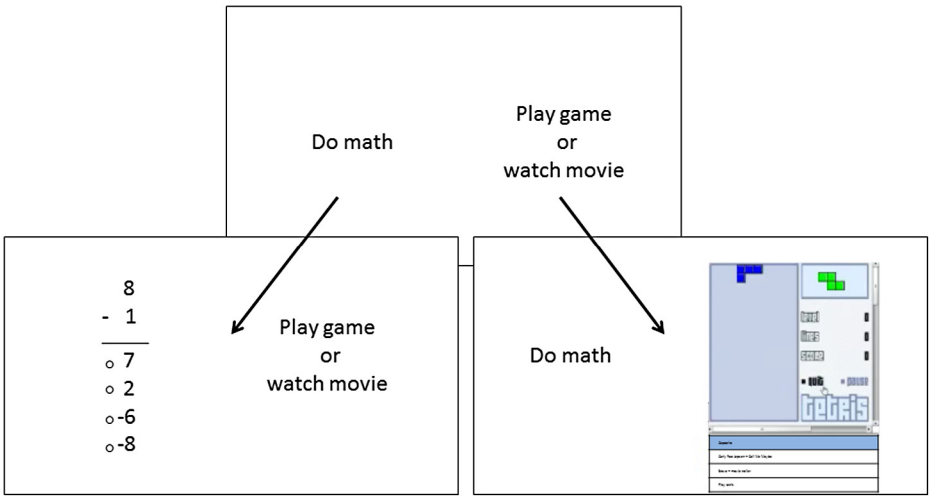

Fig. S3. In the Academic Diligence Task (1), students choose between "Do math" or "Play game or watch movie." If they click "Do math," they solve single-digit subtraction problems. If they instead click "Play game or watch movie," a pull-down menu is displayed that contains various video clips or the option to play the video game Tetris. At any point the students can toggle between math or entertainment, but the program restricts engagement to one activity at a time. Figure reproduced with permission from Galla et al. (1).

Table S17. Means, standard deviations, and correlation coefficients in Study 3.

	1	2	3	4	5	6	7
1. Grit		0.59***	0.09**	0.08**	0.09**	0.27***	−0.01
2. Self-control	0.61***		0.17***	0.12***	0.03	0.32***	−0.06*
3. ADT	0.06	0.12***		0.14***	0.27***	0.25***	0.20***
4. College graduation	0.03	0.05	0.18***		0.19***	0.38***	0.04
5. SAT score	−0.06	−0.12***	0.28***	0.27***		0.47***	0.44***
6. GPA	0.27***	0.31***	0.27***	0.33***	0.31***		0.18***
7. Cognitive ability	−0.10***	−0.16***	0.24***	0.11***	0.53***	0.16***	
M	3.69	3.50	10.47	33.49%	921.11	2.91	63.10%
SD	0.74	0.70	4.95		184.24	0.67	
N	1,166	1,166	803	1,278	970	944	1,123

Notes. Correlations controlling for school dummy variables shown above the diagonal.
† p < .10, * p < .05, ** p < .01, *** p < .001.

Table S18. Multilevel logistic regression models predicting college graduation using the full sample in Study 3

	(1)	(2)	(1)	(2)	(1)	(2)
Grit: level 1 (student)	0.163*	0.151†				
	(0.070)	(0.083)				
Grit: level 2 (school)	−0.385**	−0.343*				
	(0.135)	(0.136)				
Self-control: level 1 (student)			0.161*	0.148†		
			(0.070)	(0.082)		
Self-control: level 2 (school)			−0.440***	−0.437***		
			(0.128)	(0.117)		
Academic Diligence Task: level 1 (student)					0.153*	0.138†
					(0.070)	(0.083)
Academic Diligence Task: level 2 (school)					0.460***	0.324*
					(0.128)	(0.135)
Missing Academic Diligence Task					−0.170	−0.020
					(0.158)	(0.181)
General cognitive ability	0.080	0.064	0.074	0.052	0.086	0.077
	(0.074)	(0.085)	(0.073)	(0.084)	(0.074)	(0.085)
Student demographics						
Age		−0.320*		−0.357*		−0.314*
		(0.151)		(0.149)		(0.151)
Female		0.559***		0.540**		0.583***
		(0.165)		(0.165)		(0.165)
Eligible for free or reduced-priced meals		−0.703*		−0.791**		−0.651*
		(0.300)		(0.302)		(0.301)
English Language Learner		0.417		0.573†		0.205
		(0.370)		(0.319)		(0.375)
Special Education Student		0.079		0.056		0.165
		(0.446)		(0.447)		(0.447)
Race/ethnicity						
White		−0.060		−0.110		−0.019
		(0.514)		(0.510)		(0.514)
Hispanic		0.015		−0.129		0.193
		(0.249)		(0.243)		(0.238)
Asian		0.196		0.180		0.246
		(0.375)		(0.369)		(0.372)
Other		−0.539		−0.554		−0.486
		(0.405)		(0.406)		(0.402)
Intercept	−0.750***	5.391*	−0.740***	6.194*	−0.661***	5.159†
	(0.133)	(2.712)	(0.123)	(2.670)	(0.126)	(2.706)
N	1,037	780	1,037	780	1,037	780
Log Likelihood	-639.542	-478.316	-638.410	-476.401	-637.526	-478.705
Akaike Inf. Crit.	1,289.083	984.633	1,286.821	980.802	1,287.053	987.410
Bayesian Inf. Crit.	1,313.804	1,049.863	1,311.541	1,046.032	1,316.717	1,057.299

Notes. For each variable we report 2 models. (1) Unadjusted, (2) Controlling for student characteristics. Level 2 is the empirical bayesian estimate of the school mean. Level 1 is the deviation of each individual relative to the estimated school mean. Numeric variables were standardized prior to estimation, so coefficients are betas. Categorical variables are not standardized.
† < p .10, * p < .05, ** p < .01, *** p < .001.

Table S19. Multilevel logistic regression models predicting college graduation using listwise deletion in Study 3

	(1)	(2)	(1)	(2)	(1)	(2)
Grit: level 1 (student)	0.163*	0.151†				
	(0.070)	(0.083)				
Grit: level 2 (school)	−0.385**	−0.343*				
	(0.135)	(0.136)				
Self-control: level 1 (student)			0.161*	0.148†		
			(0.070)	(0.082)		
Self-control: level 2 (school)			−0.440***	−0.437***		
			(0.128)	(0.117)		
Academic Diligence Task: level 1 (student)					0.154*	0.138†
					(0.070)	(0.083)
Academic Diligence Task: level 2 (school)					0.461***	0.324*
					(0.129)	(0.136)
General cognitive ability					0.094	0.078
					(0.073)	(0.085)
Student demographics						
Age		−0.320*		−0.357*		−0.315*
		(0.151)		(0.149)		(0.151)
Female		0.559***		0.540**		0.585***
		(0.165)		(0.165)		(0.165)
Eligible for free or reduced-priced meals		−0.703*		−0.791**		−0.653*
		(0.300)		(0.302)		(0.300)
English Language Learner		0.417		0.573†		0.203
		(0.370)		(0.319)		(0.376)
Special Education Student		0.079		0.056		0.162
		(0.446)		(0.447)		(0.446)
Race/ethnicity						
White		−0.060		−0.110		−0.015
		(0.514)		(0.510)		(0.513)
Hispanic		0.015		−0.129		0.193
		(0.249)		(0.243)		(0.238)
Asian		0.196		0.180		0.249
		(0.375)		(0.369)		(0.371)
Other		−0.539		−0.554		−0.486
		(0.405)		(0.406)		(0.402)
Intercept	−0.750***	5.391*	−0.740***	6.194*	−0.708***	5.174†
	(0.133)	(2.712)	(0.123)	(2.670)	(0.119)	(2.704)
N	1,037	780	1,037	780	1,037	780
Log Likelihood	-639.542	-478.316	-638.410	-476.401	-638.103	-478.711
Akaike Inf. Crit.	1,289.083	984.633	1,286.821	980.802	1,286.205	985.422
Bayesian Inf. Crit.	1,313.804	1,049.863	1,311.541	1,046.032	1,310.926	1,050.652

Notes. For each variable we report 2 models. (1) Unadjusted, (2) Controlling for student characteristics. Level 2 is the empirical bayesian estimate of the school mean. Level 1 is the deviation of each individual relative to the estimated school mean. Numeric variables were standardized prior to estimation, so coefficients are betas. Categorical variables are not standardized.

† < p .10, * p < .05, ** p < .01, *** p < .001.

Table S20. OLS regression models predicting self-reported grit and self-control, and the Academic Diligence Task from peer and own performance in Study 3 (conceptual replication of Studies 1 and 2).

	ADT				Grit				Self-Control			
	CG	CA	GPA	SAT	CG	CA	GPA	SAT	CG	CA	GPA	SAT
Own performance	0.117	0.204**	0.182**	0.286***	0.073	−0.018	0.315***	0.079	0.095***	−0.060	0.328***	0.027
	(0.058)	(0.046)	(0.072)	(0.073)	(0.058)	(0.037)	(0.064)	(0.075)	(0.018)	(0.037)	(0.013)	(0.048)
Peer performance	0.016	0.078**	0.023	−0.110	−0.212***	−0.214***	−0.005	−0.260***	−0.239***	−0.243***	−0.066***	−0.230***
	(0.053)	(0.024)	(0.072)	(0.08)	(0.041)	(0.04)	(0.038)	(0.054)	(0.028)	(0.047)	(0.017)	(0.021)
Student demographics												
Eligible for free or reduced-price meals	−0.026	0.064	−0.055	0.010	0.050	0.099	0.258*	0.109	0.053	0.033	0.305	0.054
	(0.152)	(0.237)	(0.116)	(0.157)	(0.065)	(0.141)	(0.104)	(0.151)	(0.184)	(0.09)	(0.155)	(0.585)
Female	0.096	0.138	0.032	0.145	−0.048	−0.013	−0.114	−0.007	0.200***	0.268***	0.097	0.270***
	(0.117)	(0.116)	(0.16)	(0.083)	(0.047)	(0.065)	(0.055)	(0.088)	(0.034)	(0.045)	(0.051)	(0.043)
English language learner	0.338	0.379*	0.373	0.062	−0.041	−0.229	−0.297	−0.689*	0.264**	0.074	0.051	−0.261
	(0.189)	(0.159)	(0.32)	(0.663)	(0.321)	(0.242)	(0.223)	(0.309)	(0.092)	(0.081)	(0.082)	(0.134)
Special education student	−0.491*	−0.490*	−0.220	−0.079	0.221	0.325***	0.431	0.296**	0.226	0.172	0.486***	0.161
	(0.222)	(0.21)	(0.227)	(0.305)	(0.118)	(0.091)	(0.237)	(0.077)	(0.106)	(0.165)	(0.135)	(0.148)
Race/ethnicity												
White	0.198	0.177	0.153	0.129	0.087	0.079	−0.046	0.074	−0.038	−0.022	−0.019	−0.139
	(0.133)	(0.183)	(0.091)	(0.191)	(0.225)	(0.334)	(0.077)	(0.28)	(0.312)	(0.431)	(0.131)	(0.787)
Hispanic	0.229*	0.011	0.246	0.203*	−0.071	0.047	−0.083	−0.066	−0.190	−0.071	−0.267**	−0.191***
	(0.097)	(0.099)	(0.14)	(0.085)	(0.164)	(0.097)	(0.14)	(0.11)	(0.121)	(0.069)	(0.083)	(0.039)
Asian	0.645***	0.532***	0.604***	0.519***	0.194	0.227	−0.121	0.177	0.077	0.149	−0.220	−0.214
	(0.117)	(0.089)	(0.09)	(0.118)	(0.446)	(0.22)	(0.365)	(0.114)	(0.123)	(0.182)	(0.143)	(0.158)
Other	0.054	0.009	0.145	0.031	0.161	0.133	0.054	0.209	−0.099	−0.174	−0.188	−0.216
	(0.224)	(0.2)	(0.287)	(0.22)	(0.192)	(0.195)	(0.253)	(0.227)	(0.166)	(0.216)	(0.188)	(0.222)
Intercept	−0.218	−0.211	−0.211	−0.293	0.000	−0.134	−0.035	−0.074	−0.112	−0.195	−0.155	−0.149
	(0.332)	(0.393)	(0.308)	(0.23)	(0.084)	(0.162)	(0.105)	(0.189)	(0.174)	(0.099)	(0.152)	(0.462)

Notes. ADT = Academic Diligence Task, CG = College graduation, CA = Cognitive ability. Standard errors were calculated from the 95% bootstrap confidence intervals. They are calculated under the assumption of symmetric confidence intervals that are t-distributed. These assumptions do not necessarily hold in the wild bootstrap procedure.
† p < .10, * p < .05, ** p < .01, *** p < .001.

SI Dataset S1 (S2-CleanData.dta)
Dataset for Study 2.

SI Dataset S2 (S3-CleanData.rda)
Dataset for Study 3.

References

1. BM Galla, et al., The Academic Diligence Task (ADT): Assessing individual differences in effort on tedious but important schoolwork. *Contemp. Educ. Psychol.* **39**, 314–325 (2014).